

Bruchpunktschätzung bei der Ratingklassenbildung

Dissertation
zur Erlangung des akademischen Grades
doctor rerum politicarum (Dr. rer. pol.)

vorgelegt an der
Fakultät Wirtschaftswissenschaften
der Technischen Universität Dresden

von
Dipl.-Math. Daniel Tillich

Englischer Titel: Rating Classification via Split-Point Estimation

Tag der Einreichung: 18.02.2013

Tag der Disputation: 09.07.2013

Gutachter: Prof. Dr. Stefan Huschens

 Prof. Dr. Hermann Locarek-Junge

Druck: Optimiert für beidseitigen Druck.

Für Bastian

Ich danke ...

Ich danke an erster Stelle meinem Doktorvater und Erstgutachter, Herrn Prof. Dr. Stefan Huschens, der während meiner Tätigkeit am Lehrstuhl für Quantitative Verfahren, insbesondere Statistik der Technischen Universität Dresden und speziell während der Erstellung der vorliegenden Arbeit Begleiter, Kritiker und Ratgeber war. Er war es zudem, der mich in das Gebiet der Kreditrisikomodellierung einführte.

Ich danke Herrn Prof. Dr. Hermann Locarek-Junge für die Anfertigung des Zweitgutachtens zu dieser Dissertation.

Ich danke Herrn Prof. Dr. Dietmar Ferger dafür, dass er mich schon zu Studienzeiten durch spannende, forschungsnahe Vorlesungen auf das Gebiet der Statistik und besonders auf die Bruchpunktschätzung aufmerksam gemacht hat.

Ich danke meinen Kollegen, Frau PD Dr. Steffi Höse, Herrn Dr. Christoph Lehmann und Frau Sigrid Masurat, für die gute Zusammenarbeit, die wissenschaftliche Diskussion und so manchen Hinweis, der zur Lösung technischer Probleme führte.

Ich danke meiner Familie. Hier gilt mein Dank zuerst meinen Eltern für ihre unglaubliche Unterstützung in allen Lebenslagen. Des Weiteren danke ich meinen Schwestern dafür, dass sie mich früher oder später („Als Anregung zum Nach-, Besser- oder auch Gar-Nicht-Machen – ganz wie Du magst!“) auf den Weg der Mathematik und der Wissenschaft gebracht haben.

Ich danke meiner Frau Anne: „Danke, dass Du mir den Rücken freigehalten, stets Verständnis entgegengebracht und, wenn ich es brauchte, Mut zugesprochen hast.“

Dresden, Dezember 2013

Daniel Tillich

Kurzzusammenfassung

Ratingsysteme sind ein zentraler Bestandteil der Kreditrisikomodellierung. Neben der Bonitätsbeurteilung auf der Ebene der Kreditnehmer und der Risikoquantifizierung auf der Ebene der Ratingklassen spielt dabei die Bildung der Ratingklassen eine wesentliche Rolle. Die Literatur zur Ratingklassenbildung setzt auf modellfreie, in gewisser Weise willkürliche Optimierungsverfahren. Ein Ziel der vorliegenden Arbeit ist es, stattdessen ein parametrisches statistisches Modell zur Bildung der Ratingklassen einzuführen. Ein geeignetes Modell ist im Bereich der Bruchpunktschätzung zu finden. Dieses Modell und die in der mathematischen Literatur vorgeschlagenen Parameter- und Intervallschätzer werden in der vorliegenden Arbeit dargestellt und gründlich diskutiert. Dabei wird Wert auf eine anwendungsnahe und anschauliche Formulierung der mathematisch-statistischen Sachverhalte gelegt. Anschließend wird die Methodik der Bruchpunktschätzung auf einen konkreten Datensatz angewendet und mit verschiedenen anderen Kriterien zur Ratingklassenbildung verglichen. Hier erweist sich die Bruchpunktschätzung als vorteilhaft. Aufbauend auf der empirischen Untersuchung wird abschließend weiterer Forschungsbedarf abgeleitet. Dazu werden insbesondere Konzepte für den Mehrklassenfall und für abhängige Daten entworfen.

Abstract

Rating systems are a key component of credit risk modeling. In addition to scoring at borrowers' level and risk quantification at the level of rating classes, the formation of the rating classes plays a fundamental role. The literature on rating classification uses in a way arbitrary optimization methods. Therefore, one aim of this contribution is to introduce a parametric statistical model to form the rating classes. A suitable model can be found in the area of split-point estimation. This model and the proposed parameter and interval estimators are presented and thoroughly discussed. Here, emphasis is placed on an application-oriented and intuitive formulation of the mathematical and statistical issues. Subsequently, the methodology of split-point estimation is applied to a specific data set and compared with several other criteria for rating classification. Here, split-point estimation proves to be advantageous. Finally, further research questions are derived on the basis of the empirical study. In particular, concepts for the case of more than two classes and for dependent data are sketched.

Inhaltsverzeichnis

Abbildungsverzeichnis	VII
Tabellenverzeichnis	IX
Abkürzungsverzeichnis	XI
Häufig verwendete Symbole	XIII
1 Einleitung	1
1.1 Problemstellung, Ziele, Abgrenzung	1
1.2 Aufbau der Arbeit	3
2 Ratingsystem	5
2.1 Bonitätsbeurteilung auf Kreditnehmerebene	7
2.1.1 Scoringmodelle	7
2.1.2 Vergleich von Scoringssystemen	9
2.2 Ratingklassenbildung	18
2.2.1 Methoden	20
2.2.2 Vergleichskriterien	26
2.3 Risikoquantifizierung auf Ratingklassenebene	29
3 Bruchpunktschätzung	31
3.1 Modell und Schätzproblem	32
3.1.1 Modellstruktur	32
3.1.2 Eigenschaften des Modells	36
3.1.3 Schätzproblem	38
3.2 Dempfle-Stute-Schätzer	39
3.2.1 Einführung	39
3.2.2 Äquivalente Schreibweisen	43
3.2.3 Verwandte Schätzer und asymptotische Eigenschaften	51
3.3 Maximum-Likelihood-Schätzer bei bekannten Risikoniveaus	58
3.4 Plug-In-Schätzer bei unbekannten Risikoniveaus	67

3.4.1	Schätzung der Risikoniveaus	68
3.4.2	Schätzung des Bruchpunkts	73
3.4.3	Simultane Schätzung	78
3.5	Vergleich der Schätzer	79
4	Anwendung	83
4.1	Datensatz und Modellannahmen	83
4.1.1	Einführung	83
4.1.2	Univariate Analysen	86
4.1.3	Multivariate Analysen	91
4.2	Bruchpunktschätzung für einzelne Quartale	101
4.2.1	Maximum-Likelihood-Schätzer	102
4.2.2	Dempfle-Stute- und Plug-In-Schätzer	110
4.2.3	Konfidenzintervalle und Konfidenzregionen	120
4.3	Bruchpunktschätzung im Pool aller Quartale	130
5	Zusammenfassung und Ausblick	135
5.1	Zusammenfassung	135
5.2	Ausblick	137
	Literaturverzeichnis	145

Abbildungsverzeichnis

2.1	Empirische Verteilungsfunktionen in den Beispieldatensätzen	12
2.2	CAP- und ROC-Kurven der Beispieldatensätze	14
2.3	Empirischer Prozess d^{ROC} für Datensätze A und B	18
3.1	Funktion κ für $a > b$	33
3.2	Empirischer Prozess S_n für Datensätze A und B	44
3.3	Vier Realisationen des Zufallsvektors (X, Y)	48
3.4	Realisierter Pfad von S_n	48
4.1	Anzahl der Beobachtungen im Zeitablauf	84
4.2	Absolute Häufigkeit der Bonitätsscores	86
4.3	Empirische Verteilungsfunktion der Bonitätsscores	87
4.4	Kerndichtefunktionen der Bonitätsscores	89
4.5	Anzahl der Ausfälle im Zeitablauf	90
4.6	Ausfallquoten im Zeitablauf	91
4.7	Absolute Häufigkeit der Scores der Ausfälle	92
4.8	Empirische Verteilungsfunktion der Scores der Ausfälle	92
4.9	Kerndichtefunktionen für die Scores der Ausfälle	93
4.10	Empirische Regression 1. Art	95
4.11	Empirische Regression 1. Art, Log-Skala	96
4.12	Kernregression	98
4.13	ROC-Kurven für die Quartale 1, 21 und 43	99
4.14	Funktion d^{ROC} für die Quartale 1, 21 und 43	100
4.15	ROC-gap (=KS-Statistik) im Zeitablauf	100
4.16	Funktionsverlauf des empirischen Prozesses S_n^* , drei Quartale	103
4.17	Funktionsverlauf des empirischen Prozesses S_n^* , vier Parameterpaare	103
4.18	Bruchpunktschätzwerte im Zeitablauf, bekannte Risikoniveaus	106
4.19	CH-Statistik für die Bruchpunktschätzwerte im Zeitablauf, bekannte Risikoniveaus	108
4.20	Funktionsverlauf des empirischen Prozesses S_n	112
4.21	Bruchpunktschätzwerte nach Dempfle & Stute im Zeitablauf	113

4.22	Zeitreihen der (Teil-)Ausfallquoten	114
4.23	Funktionsverlauf des empirischen Prozesses \bar{S}_n^*	115
4.24	Bruchpunktschätzwerte im Zeitablauf, unbekannte Risikoniveaus	116
4.25	CH-Statistik für die Bruchpunktschätzwerte im Zeitablauf, unbekannte Risikoniveaus	118
4.26	Aus Partialsummen zusammengesetzter Prozess	124
4.27	Abweichungen der Konfidenzintervallgrenzen von den jeweiligen Bruch- punktschätzwerten im Zeitablauf, bekannte Risikoniveaus	127
4.28	Abweichungen der Intervallgrenzen der simultanen Konfidenzregionen von den jeweiligen Schätzwerten im Zeitablauf, unbekannte Risikoniveaus	129
4.29	Funktionsverlauf der empirischen Prozesse im Pool aller Quartale . . .	131
4.30	Zeitreihen der auf den Pool-Schätzwerten basierenden quartalsweisen Teilausfallquoten	133

Tabellenverzeichnis

2.1	Vergleich von prognostiziertem und beobachtetem Ausfallverhalten . . .	11
3.1	Denkbare und untersuchte Anwendungsbeispiele des Modells	35
4.1	Unbesetzte Scores in drei ausgewählten Quartalen	94
4.2	Lage- und Streuungsmaßzahlen zu den Bruchpunktschätzwerten bei bekannten Risikoniveaus	106
4.3	Empirische Korrelationsmatrix der Bruchpunktschätzwerte und der Ausfallquote, bekannte Risikoniveaus	107
4.4	Deskriptive Statistik zur CH-Statistik, bekannte Risikoniveaus	108
4.5	Deskriptive Statistik zum Anteil der Ratingklasse 1, bekannte Risikoniveaus	109
4.6	Lage- und Streuungsmaßzahlen sowie empirische Korrelationen zu den (Teil-)Ausfallquoten	114
4.7	Lage- und Streuungsmaßzahlen zu den Bruchpunktschätzwerten bei unbekannten Risikoniveaus	116
4.8	Empirische Korrelationsmatrix zu den Bruchpunktschätzwerten bei unbekannten Risikoniveaus, zum Maximum der Calinski-Harabasz-Statistik und zur Ausfallquote im Zeitablauf	117
4.9	Deskriptive Statistik zur CH-Statistik, unbekannte Risikoniveaus	118
4.10	Deskriptive Statistik zum Anteil der Ratingklasse 1, unbekannte Risikoniveaus	119
4.11	Bei der Quantilbestimmung interessierende Wahrscheinlichkeiten	122
4.12	Lage- und Streuungsmaßzahlen zu den geschätzten Intensitäten des zusammengesetzten Poisson-Prozesses	123
4.13	Lage- und Streuungsmaßzahlen der durch Simulation ermittelten kleinsten und größten Maximalstelle τ bzw. σ des zusammengesetzten Poisson-Prozesses (Werte gerundet), verschiedene Eingangsparameter, unterschiedliche Anzahl \mathcal{N} von Monte-Carlo-Replikationen	125

4.14	Konfidenzintervalle für θ bei bekannten Risikoniveaus $a = 0.01$ und $b = 0.001$ für drei ausgewählte Quartale und zwei verschiedene Konfidenzniveaus	126
4.15	Konfidenzregionen für (θ, a, b) bei unbekannten Risikoniveaus für drei ausgewählte Quartale und zwei verschiedene Konfidenzniveaus	128

Abkürzungsverzeichnis

arithm.	arithmetisch(es)
B	Referenzskala (<i>benchmark</i>)
BCBS	Basel Committee on Banking Supervision
bzw.	beziehungsweise
CAP	<i>Cumulative Accuracy Profile</i>
CH	Calinski-Harabasz
CMT	Stetigkeitssatz (<i>continuous mapping theorem</i>)
D	Ausfall (<i>default</i>)
d. h.	das heißt
DSGV	Deutscher Sparkassen- und Giroverband
EAD	Forderungshöhe bei Ausfall (<i>exposure at default</i>)
FAR	<i>False Alarm Rate</i>
f. s.	fast sicher
HR	<i>Hit Rate</i>
i.i.d.	stochastisch unabhängig und identisch verteilt (<i>independent and identically distributed</i>)
insb.	insbesondere
IRB(A)	<i>Internal Ratings-Based (Approach)</i>
KD	Kerndichte
KS	Kolmogoroff-Smirnoff
lcrl	linksseitig stetig mit rechtsseitigen Grenzwerten (<i>left continuous with right-hand limits</i>)
LGD	Verlustquote bei Ausfall (<i>loss given default</i>)
ML	Maximum-Likelihood

MSE	<i>mean squared error</i>
ND	Nichtausfall (<i>non-default</i>)
PD	Ausfallwahrscheinlichkeit (<i>probability of default</i>)
Q	Quartal
rcll	rechtsseitig stetig mit linksseitigen Grenzwerten (<i>right continuous with left-hand limits</i>)
RMSE	<i>root mean squared error</i>
ROC	<i>Receiver Operating Characteristic</i>
S.	Seite(n)
Standardabw.	Standardabweichung
st. u.	stochastisch unabhängig
TTC	<i>Trough-the-Cycle</i>
u.	und
vgl.	vergleiche
Z.	Zeile(n)
z. B.	zum Beispiel

Häufig verwendete Symbole

a	Risikoniveau, Funktionswert von κ
a_n^*	Maximum-Likelihood-Schätzer für a
\bar{a}_n^*	Plug-In-Schätzer für a
argmax	Maximalstelle
argmin	Minimalstelle
b	Risikoniveau, Funktionswert von κ
b_n^*	Maximum-Likelihood-Schätzer für b
\bar{b}_n^*	Plug-In-Schätzer für b
$B(m, p)$	Binomialverteilung mit Parametern m und p
$B_r(\theta)$	r -Umgebung von θ
$\operatorname{Ber}(p)$	Bernoulli-Verteilung mit Parameter p
$C_n(\gamma)$	asymptotisches Konfidenzintervall für (θ, a, b) zum Konfidenzniveau $1 - \gamma$
CH	Calinski-Harabasz-Statistik
Cov	Kovarianz
d^{ROC}	Differenz von Hit Rate und False Alarm Rate
\mathbb{E}	Erwartungswert
\exp	Exponentialfunktion
f	Dichtefunktion der Scores
f_n	Schätzer für die Dichtefunktion der Scores
\hat{f}_h	Kerndichteschätzer für die Dichtefunktion der Scores
$\hat{f}_h^{\operatorname{D}}$	Kerndichteschätzer für die Dichtefunktion der Scores der Ausfälle (<i>defaults</i>)
F	Verteilungsfunktion der Scores
F^{D}	Verteilungsfunktion der Scores der Ausfälle (<i>defaults</i>)

F^{ND}	Verteilungsfunktion der Scores der Nichtausfälle (<i>non-defaults</i>)
F_n	empirische Verteilungsfunktion der Scores
F_n^{\leftarrow}	empirische Quantilfunktion zu F
FAR	empirische Verteilungsfunktion der Scores der Nichtausfälle
h, h_n	Bandbreite bei Kerndichteschätzung oder Kernregression
H_n	markierte empirische Verteilungsfunktion der Scores
HR	empirische Verteilungsfunktion der Scores der Ausfälle
$I_n(\gamma)$	asymptotisches Konfidenzintervall für θ zum Konfidenzniveau $1 - \gamma$
K	Anzahl der Ratingklassen
\mathcal{K}	Kernfunktion
L	zusammengesetzter Poisson-Prozess
\mathcal{L}_n	Likelihoodfunktion
ℓ_n	Loglikelihoodfunktion
\liminf	Limes inferior
\limsup	Limes superior
\ln	natürlicher Logarithmus
\max	Maximum
\min	Minimum
n	Anzahl der Beobachtungen (zum Teil versehen mit Indizes zur Angabe von Periode t oder Ratingklasse k)
\mathbb{N}, \mathbb{N}_0	Menge der natürlichen Zahlen (inklusive der Null)
$N(\mu, \sigma^2)$	Normalverteilung mit Parametern μ und σ^2
$N_{\text{D}}, n_{\text{D}}$	(zufällige) Anzahl der Ausfälle
N_{ND}	zufällige Anzahl der Nichtausfälle
\mathcal{N}	Anzahl der Monte-Carlo-Wiederholungen
\mathcal{O}_P	Ordnung in Wahrscheinlichkeit
$P(E)$	Wahrscheinlichkeit des Ereignisses E
p_k	Risikoniveau der k -ten Klasse
\mathbb{R}	Menge der reellen Zahlen
r_n	empirischer Prozess, $r_n(t) := S_n(F_n^{\leftarrow}(t))$

ROC-gap	maximale Differenz von Hit Rate und False Alarm Rate
S	Grenzprozess zu S_n
S_n	empirischer Prozess, $S_n(x) := H_n(x) - \bar{Y}F_n(x)$
S_n^*	empirischer Prozess, $S_n^*(x) := \alpha H_n(x) + \beta F_n(x)$
\bar{S}_n^*	empirischer Prozess, $\bar{S}_n^*(x) := \bar{\alpha}_n^* H_n(x) + \bar{\beta}_n^* F_n(x)$
sup	Supremum
T	Anzahl der Perioden
V	Varianz
V_k	Partialsummenprozess
W_m	Partialsummenprozess
X	(Bonitäts-)Score (zum Teil versehen mit Indizes zur Angabe von Beobachtung i oder Periode t)
$X_{i:n}$	i -te Variable der aufsteigend geordneten Stichprobe
\tilde{X}_p, \tilde{x}_p	(zufälliges) empirisches p -Quantil der Verteilung von X
Y	Ausfallindikator (zum Teil versehen mit Indizes zur Angabe von Beobachtung i , Periode t oder Ratingklasse k)
\bar{Y}	Ausfallquote (zum Teil versehen mit Indizes zur Angabe von Periode t oder Ratingklasse k)
y_k^B	Ausfallwahrscheinlichkeit der k -ten Ratingklasse der Referenzskala (<i>benchmark</i>)
\mathbb{Z}	Menge der ganzen Zahlen
α	$\alpha := \ln\left(\frac{a}{b} \frac{1-b}{1-a}\right)$
$\bar{\alpha}_n^*$	Plug-In-Schätzer für α
β	$\beta := \ln\left(\frac{1-a}{1-b}\right)$
$\bar{\beta}_n^*$	Plug-In-Schätzer für β
β_k	Regressionskoeffizienten der logistischen Regression
$\hat{\beta}_k$	Schätzwerte für die Regressionskoeffizienten der logistischen Regression
ϵ	Störgröße der Regression
θ	Bruchpunkt, Trennstelle der Ratingklassen
$\hat{\theta}_n$	(einseitiger) Dempfle-Stute-Schätzer für θ ($a > b$)

$\hat{\theta}_n^-$	(einseitiger) Dempfle-Stute-Schätzer für θ ($a < b$)
$\bar{\theta}_n$	zweiseitiger Dempfle-Stute-Schätzer für θ
θ_n^*	Maximum-Likelihood-Schätzer für θ
$\bar{\theta}_n^*$	(Zwei-Schritt-)Plug-In-Schätzer für θ
$\hat{\theta}_A$	Anderson-Schätzer für θ
$\hat{\theta}_{A^*}$	modifizierter Anderson-Schätzer für θ
$\hat{\theta}_C$	Konstante als Schätzer für θ
$\hat{\theta}_{CH}$	Calinski-Harabasz-Schätzer für θ
$\hat{\theta}_F$	Fernandes-Schätzer für θ
$\hat{\theta}_{F^*}$	modifizierter Fernandes-Schätzer für θ
$\hat{\theta}_M$	Median der empirischen Scoreverteilung als Schätzer für θ
κ	Regressionsfunktion
$\hat{\kappa}_h$	Nadaraya-Watson-Schätzer für κ
λ	Intensität des Poisson-Prozesses
λ_n^*	Intensitätsschätzer basierend auf θ_n^*
$\bar{\lambda}_n$	Intensitätsschätzer basierend auf $\bar{\theta}_n$
$\bar{\lambda}_n^*$	Intensitätsschätzer basierend auf $\bar{\theta}_n^*$
μ_n	Schätzer für die Maximalstelle von r_n
π_i	Ausfallwahrscheinlichkeit des i -ten Individuums
$\hat{\pi}_i$	Schätzwert für die individuenspezifische Ausfallwahrscheinlichkeit π_i
$\bar{\pi}_k$	arithmetisches Mittel der Schätzwerte $\hat{\pi}_i$ in Ratingklasse k
σ	größte Maximalstelle von L
τ	kleinste Maximalstelle von L
φ	Dichtefunktion der Standardnormalverteilung
Φ	Verteilungsfunktion der Standardnormalverteilung
Φ^{-1}	Quantilfunktion der Standardnormalverteilung, Inverse von Φ
$\mathbb{1}$	Indikatorfunktion
$:=$	definitorisches Gleichheitszeichen
\sim	verteilt nach

\xrightarrow{d}	Konvergenz in Verteilung (<i>convergence in distribution</i>)
$\xrightarrow{\text{f. s.}}$	fast sichere Konvergenz
$\rightsquigarrow_{a.w.}$	fast schwache Konvergenz (<i>almost weak convergence</i>)
$g(x-)$	linksseitiger Grenzwert der Funktion g an der Stelle x , $g(x-) := \lim_{\varepsilon \downarrow 0} g(x - \varepsilon)$
$g(x+)$	rechtsseitiger Grenzwert der Funktion g an der Stelle x , $g(x+) := \lim_{\varepsilon \downarrow 0} g(x + \varepsilon)$
\square	Ende eines Beweises

Kapitel 1

Einleitung

1.1 Problemstellung, Ziele, Abgrenzung

Während der Finanzkrise 2007/2008, aber auch in der darauf folgenden und bis heute andauernden sogenannten Staatsschuldenkrise, gerieten weltweit Banken in Schieflage. Daraufhin mussten etliche von ihnen von öffentlicher Hand massiv gestützt werden, damit die Kreditvergabe aufrecht erhalten werden konnte. Der Grund für die finanziellen Probleme der Banken war in vielen Fällen eine unzureichende Vorsorge für eingegangene Kreditrisiken. Insofern ist eine risikoadäquate Eigenkapitalunterlegung von immensem Wert.

Als Grundlage für die Berechnung der Höhe des vorzuhaltenden Eigenkapitals dienen Ratingsysteme. Diese haben die Aufgabe, insbesondere das Ausfallrisiko der Vertragspartner zu quantifizieren. Die zentrale Größe dabei ist die Ausfallwahrscheinlichkeit. Sie bezeichnet die Wahrscheinlichkeit dafür, dass innerhalb eines Zeitraums von in der Regel einem Jahr das Ausfallereignis eintritt. Das Ausfallereignis tritt ein, wenn der Kreditnehmer seine Zahlungsverpflichtungen nicht vollständig bedienen kann. Die Ausfallwahrscheinlichkeit eines potenziellen Vertragspartners der Bank kann sehr unterschiedlich sein und prinzipiell jeden Wert zwischen 0 und 1 annehmen. Man kann hier also von einer kontinuierlichen Skala ausgehen.

Im Gegensatz dazu werden innerhalb der Ratingsysteme Kreditnehmer mit ähnlichem Ausfallrisiko jedoch zu Gruppen zusammengefasst. Diesen sogenannten Ratingklassen werden anschließend Ausfallwahrscheinlichkeiten zugeordnet, so dass für alle Individuen (Personen, Unternehmen oder Staaten) in einer Klasse ein einheitliches Ausfallrisiko unterstellt wird. Durch diesen Vorgang wird das ursprünglich kontinuierliche Ausfallrisiko diskretisiert. Hierbei stellt sich die Frage, auf welcher Grundlage bzw. nach welchem Kriterium die Ratingklassen gebildet werden sollten. Beim Blick in die Literatur zeigt sich, dass dort die Trennstellen zwischen den Klassen modellfrei festgelegt werden. In den fortgeschrittenen Ansätzen erfolgt die Einteilung der Ratingklassen

anhand eines in gewisser Weise sinnvollen, in gewisser Weise jedoch auch willkürlich gewählten Optimierungskriteriums.

Ein Ziel dieser Arbeit ist es daher,

- für das Problem der Klassenbildung ein parametrisches statistisches Modell einzuführen.

Darauf aufbauend gilt es,

- geeignete Schätzer für die unbekannten Modellparameter zu finden und
- die Schätzer einer Güteuntersuchung zu unterziehen.

Es zeigte sich, dass die einfachste Variante des Modells, der Zweiklassenfall, bereits im Gebiet der Bruchpunktschätzung (*change-point analysis*) auftaucht. Von Dempfle & Stute (2002) sowie Ferger & Klotsche (2009) wurden Schätzer vorgeschlagen und bezüglich ihrer asymptotischen Eigenschaften untersucht. Die vorliegende Arbeit will

- diese Literatur kritisch durchsehen, gegebenenfalls Lücken schließen und eventuelle Verbindungen zum Kreditrisiko aufdecken.

Anschließend werden die vorgestellten Methoden erstmalig dazu verwendet, Bonitäts-scores in Ratingklassen einzuteilen. Als Grundlage dient dabei ein Datensatz einer großen deutschen Auskunft. Diese praktische Anwendung soll

- überprüfen, wie die Ansätze der Bruchpunktschätzung im Vergleich zu den sonstigen diskutierten Ideen abschneiden und welche zusätzlichen Ergebnisse sich erzielen lassen.

Hierbei handelt es sich um Konfidenzintervalle für den Bruchpunkt bzw. simultane Konfidenzbereiche für den Bruchpunkt und die ratingklassenspezifischen Ausfallwahrscheinlichkeiten. Das abschließende Ziel dieser Arbeit ist es,

- basierend auf der empirischen Untersuchung weiteren Forschungsbedarf abzuleiten und die Stoßrichtungen zukünftiger Forschung zu skizzieren.

An dieser Stelle sei sogleich angemerkt, dass es bei der vorliegenden Form der Bruchpunktschätzung darum geht, einen Schwellenwert innerhalb des Wertebereichs einer ersten Zufallsvariable zu finden, an dem sich das Verhalten einer zweiten Zufallsvariable grundlegend ändert. Nicht gemeint ist das Change-Point-Problem der Zeitreihenanalyse, bei dem nach einem Zeitpunkt gesucht wird, an dem ein Strukturbruch stattgefunden hat. Solche Regimewechsel treten z. B. bei Finanzmarktzeitreihen auf und äußern sich z. B. in einem veränderten Erwartungswert, einer veränderten Varianz oder einer veränderten (Auto-)Korrelation.¹ Eng verwandt mit der sequentiellen

¹Vgl. für dieses Problem z. B. Hamilton (1994, Kapitel 22) oder Steland (2012, Abschnitt 9.3).

Suche nach einem Bruchpunkt in der Zeit ist die statistische Qualitätskontrolle bei Produktionsprozessen mit Qualitätsregel- oder Kontrollkarten.²

Da das Modell der Bruchpunktschätzung zunächst für zwei Klassen aufgestellt wird, sollte weiterhin erwähnt werden, dass mit dem Bruchpunkt keine Prognose von Ausfall oder Nichtausfall verknüpft werden soll. Ebenfalls soll der Bruchpunkt nicht als Schwellenwert zwischen einer Ablehnung oder einer Zustimmung zur Kreditvergabe verstanden werden. Für Entscheidungen dieser Art sind immer die entsprechenden Kosten der Bank bei Fehleinschätzungen von Ausfall oder Nichtausfall zu berücksichtigen.³ In unserem Modell geht es vielmehr um eine Einteilung der Bestandskunden im Sinne von niedrigerer oder höherer Ausfallwahrscheinlichkeit.

1.2 Aufbau der Arbeit

Aus den in Abschnitt 1.1 genannten Zielen ergibt sich die folgende Struktur der Arbeit. Kapitel 2 beschäftigt sich mit den drei Basiselementen eines bankinternen Ratingsystems. Diese sind die Bonitätsbeurteilung auf Kreditnehmerebene, die anschließende Ratingklasseneinteilung und die abschließende Risikoquantifizierung auf Ratingklassenebene. In Abschnitt 2.1 werden Scoringmodelle, d. h. Modelle zur Beurteilung der individuenspezifischen Ausfallwahrscheinlichkeiten, besprochen. Dabei wird insbesondere auf das Scoringmodell eingegangen, welches dem Datensatz der Auskunft zu Grunde liegt. Zudem werden Möglichkeiten zum Vergleich von Scoringmodellen vorgestellt. Welche Anforderungen an die Ratingklassenbildung gestellt werden, wird in Abschnitt 2.2 diskutiert. Anschließend werden verschiedene Methoden vorgestellt, die bei einer vorgegebenen Anzahl von Ratingklassen die Grenzen der Klassen festlegen. Dabei wird unterschieden, ob für die Ratingklassen Zielausfallwahrscheinlichkeiten vorgegeben sind oder nicht. Ferner betrachten wir einen Ansatz, der den Vergleich verschiedener Einteilungsvarianten ermöglicht. Für den Fall, dass keine Referenzgrößen für die Ausfallwahrscheinlichkeiten der Ratingklassen vorgegeben sind, müssen abschließend die ratingklassenspezifischen Ausfallwahrscheinlichkeiten geschätzt werden. Damit beschäftigt sich Abschnitt 2.3.

In der Literatur zur Ratingklasseneinteilung fanden Methoden aus dem Bereich der Bruchpunktschätzung bisher keine Beachtung. Dabei existieren entsprechende Methoden zumindest für den Spezialfall zweier Ratingklassen und unterstellter Unabhängigkeit der Kreditnehmer. Mit diesen Methoden beschäftigt sich Kapitel 3. Hierfür werden in Abschnitt 3.1 zunächst das statistische Modell und seine Eigenschaften vorgestellt. Dann folgt eine Einführung in das Schätzproblem. Anschließend behandelt Abschnitt 3.2 den sogenannten Dempfle-Stute-Schätzer für den gesuchten Bruch-

²Vgl. z. B. Rinne (2008, Abschnitt D.6.2.2) oder Ross (2006, Kapitel 13).

³Vgl. z. B. Krämer & Bücker (2011, Abschnitt 5) und die dort genannte Literatur.

punkt, welcher in der Anwendung die Trennstelle zwischen den Ratingklassen darstellt. Der Schätzer wird ausführlicher als in der Literatur motiviert und seine verschiedenen Schreibweisen werden verknüpft. Des Weiteren werden asymptotische Eigenschaften beschrieben. In Abschnitt 3.3 werden der Maximum-Likelihood-Schätzer und seine asymptotischen Eigenschaften betrachtet. Vorausgesetzt wird hierbei, dass die gruppenspezifischen Erfolgswahrscheinlichkeiten, in der Anwendung also die ratingklassenspezifischen Ausfallwahrscheinlichkeiten, vorgegeben sind. Der aus den beiden schon genannten Ansätzen abgeleitete Substitutionsschätzer wird in Abschnitt 3.4 vorgestellt. Schließlich werden in Abschnitt 3.5 die Ergebnisse einer Simulationsstudie zum Vergleich der drei Schätzer präsentiert.

In Kapitel 4 werden die Methoden aus Kapitel 3 auf einen Auskunftsei-Datensatz angewendet. Abschnitt 4.1 stellt den Datensatz vor. Dabei wird auf die kritischen Annahmen des Modells eingegangen. Dazu werden unter anderem eine Kerndichteschätzung für die Bonitätsscores und eine Kernregression für den Zusammenhang zwischen Score und Ausfallverhalten durchgeführt. In Abschnitt 4.2 wird die Trennstelle zwischen den beiden Ratingklassen zunächst quartalsweise aus Querschnittsdaten geschätzt. Die im Zeitablauf ermittelten Bruchpunktschätzwerte werden mit den Ergebnissen ausgewählter Methoden zur Ratingklasseneinteilung aus Abschnitt 2.2 verglichen. Konfidenzintervalle für die Ratingklassengrenze sowie simultane Konfidenzregionen für die Trennstelle und die ratingklassenspezifischen Ausfallwahrscheinlichkeiten werden ebenfalls ermittelt. Anschließend wird in Abschnitt 4.3 der Bruchpunkt aus dem kompletten Datensatz, d. h. dem Pool aller Quartale, geschätzt. Eine Zusammenfassung der Ergebnisse und ein ausführlicher Ausblick auf resultierende Forschungsfragen in Kapitel 5 runden diese Arbeit ab.

Abschließend sei angemerkt,

- dass als Dezimaltrennzeichen der Dezimalpunkt verwendet wird,
- dass das Symbol \square das Ende eines Beweises markiert und
- dass in einigen Literaturverweisen die Angabe der Zeilennummer nötig ist. Dabei sind negative Zahlen so zu verstehen, dass die Zählung vom Seitenende aus erfolgt.

Kapitel 2

Ratingsystem

Die zentrale Risikokomponente des Kreditrisikos ist die Ausfallwahrscheinlichkeit (*probability of default*, PD). Sie bezeichnet die Wahrscheinlichkeit dafür, dass innerhalb eines bestimmten Zeitraums, in der Regel eines Jahres, das Ausfallereignis eintritt. Da die Ausfallwahrscheinlichkeit unbekannt ist, muss sie aus beobachteten Daten geschätzt werden.¹ Maßgeblich ist dabei auch die Definition des Ausfallereignisses. Diese kann sich von Institution zu Institution unterscheiden. Allen Definitionen des Ausfalls (*default*) gemeinsam ist, dass ein Kreditnehmer seine Zahlungsverpflichtungen nicht vollständig bedienen kann. Beispielsweise liegt gemäß den Richtlinien des Basler Ausschusses für Bankenaufsicht (*Basel Committee on Banking Supervision*, BCBS) ein Ausfall vor, wenn mindestens eines der folgenden Ereignisse eingetreten ist:²

- (i) Die Bank hält es für unwahrscheinlich, dass der Schuldner seinen Kreditverpflichtungen gegenüber der Bank in vollem Maße nachkommt, ohne dass die Bank auf Maßnahmen wie die Verwertung von Sicherheiten zurückgreifen muss.
- (ii) Der Schuldner ist mit einer wesentlichen Kreditverpflichtung gegenüber der Bank seit mehr als 90 Tagen überfällig.

Neben der Ausfallwahrscheinlichkeit definiert der Basler Ausschuss für Bankenaufsicht in seinen unter dem Namen „Basel II“ bekannten Richtlinien folgende weitere Risikokomponenten:

- Forderungshöhe bei Ausfall (*exposure at default*, EAD),
- Verlustquote bei Ausfall (*loss given default*, LGD) und
- effektive Laufzeit (*effective maturity*)

¹Vgl. z. B. Höse & Huschens (2003, S. 140).

²Vgl. BCBS (2006, S. 100, Paragraph 452) oder die Übersetzung in Höse (2007, S. 8).

Diese sollen in dieser Arbeit aber nicht Gegenstand der Betrachtung sein, denn der später verwendete Datensatz gibt zu den letztgenannten Risikokomponenten keine Auskünfte. Details zur Definition, zur (gemeinsamen) Modellierung und zur Schätzung bzw. Berechnung der Risikokomponenten lassen sich in den entsprechenden Paragraphen in BCBS (2006) sowie in der einschlägigen Literatur nachlesen.³

Wir konzentrieren uns im Folgenden auf die Ausfallwahrscheinlichkeit. Sie ist eine Eingangsgröße bei der Berechnung der Kapitalunterlegung einer Bank. Dabei lässt der Basler Ausschuss zwei Ansätze zur Bestimmung des minimalen Kapitalbedarfs für das Kreditrisiko zu:

- den Standardansatz und
- den sogenannten *Internal Ratings-Based Approach* (IRB-Ansatz).

Im IRB-Ansatz können Banken Schätzwerte aus den Daten des eigenen Ratingsystems verwenden,⁴ welches ohnehin für die bankinterne Steuerung benötigt wird.

Der Basler Ausschuss für Bankenaufsicht definiert den Begriff Ratingsystem dabei wie folgt:⁵

„The term ‚rating system‘ comprises all of the methods, processes, controls, and data collection and IT systems that support the assessment of credit risk, the assignment of internal risk ratings, and the quantification of default and loss estimates.“

Trotz der unzähligen Ausgestaltungsmöglichkeiten lässt sich die Struktur eines bankinternen Ratingsystems damit auf drei Basiselemente reduzieren.⁶

1. Bonitätsbeurteilung auf Kreditnehmerebene: Jedem Kreditnehmer wird ein Indikator für seine Kreditqualität (Bonität) zugeordnet.
2. Ratingklassenbildung: Die bewerteten Kreditnehmer werden auf Ratingklassen aufgeteilt.
3. Risikoquantifizierung auf Ratingklassenebene: Das Risiko muss in allen vorhandenen Klassen geschätzt werden.

Diese drei Basiselemente sollen in den folgenden drei Abschnitten näher beleuchtet werden. Dabei wird verstärkt auf Sachverhalte eingegangen, die im weiteren Verlauf der vorliegenden Arbeit von besonderer Bedeutung sind.

³Vgl. einführend z. B. Höse (2007, S. 9-10) oder Bluhm et al. (2010, S. 15-22).

⁴Vgl. BCBS (2006, S. 19, Paragraphen 50,51).

⁵BCBS (2006, S. 90, Paragraph 394)

⁶Vgl. auch Foglia et al. (2001, S. 422).

2.1 Bonitätsbeurteilung auf Kreditnehmerebene

Die Bonitätsbeurteilung, d. h. die Einschätzung der Kreditwürdigkeit, ist das erste Basiselement eines bankinternen Ratingsystems. Das Ziel hierbei ist es, die individuelle Ausfallwahrscheinlichkeit eines jeden Kreditnehmers zu schätzen. Auf Basis dieser Schätzung wird jedem Kreditnehmer ein Bonitätsscore oder kurz Score zugeordnet. Üblicherweise steht dabei ein hoher Score für geringes Ausfallrisiko und daher gute Bonität. Verschiedene Möglichkeiten der Scorebestimmung werden einführend in Abschnitt 2.1.1 besprochen. Mit dem Vergleich konkurrierender Scoringmodelle beschäftigt sich anschließend Abschnitt 2.1.2.

2.1.1 Scoringmodelle

In der Kreditrisikomessung haben sich automatisierte Bonitätseinschätzungen durchgesetzt, weil sie unter anderem schnell durchführbar und dabei in dem Sinne objektiv sind, dass gleiche Informationen über verschiedene Kreditnehmer auch zum gleichen Score führen. Ferner können Scoringmodelle z. B. im Hinblick auf den Zeithorizont der Ausfalldefinition an die von der Bank verwendeten Risikomodelle für das gesamte Kreditportfolio angepasst werden. Dass bei der Etablierung eines Scoringystems eine hinreichend große Anzahl von elektronisch zu verarbeitenden Daten über relevante Kreditnehmer vorliegen muss, ist heute unproblematisch.⁷

Die wohl am häufigsten verwendete Methode zur Beurteilung der Bonität von Individuen ist die logistische Regression. Sie wird auch als Logit-Regression bezeichnet. Neben ihr existiert eine Vielzahl an weiteren Scoringmodellen. Einen kurzen Überblick über die Scoring-Literatur von den Anfängen (Diskriminanzanalyse) über Logit- und Probit-Modelle hin zu alternativen nichtparametrischen Modellen (Vergleichsstudien) und aktuellen Forschungsfragen (betrifft sogenannte *rejected data*) liefert Fernandes (2005, S. 2-3). Eine Übersicht über „Marktgängige Modelle zur Bonitätsbeurteilung“ findet man in Oesterreichische Nationalbank & Finanzmarktaufsicht (2004, Kap. 3, S. 32-55). Dort wird auch kurz auf die Vorteile der Logit-Regression gegenüber der Diskriminanzanalyse und gegenüber künstlichen neuronalen Netzen eingegangen. Vertiefende Literaturhinweise fehlen jedoch größtenteils. Franke et al. (2004, S. 389-395) widmen sich in einem kurzen Kapitel ebenfalls Modellen für das Kreditscoring.

Im späteren Verlauf der Arbeit wird ein Datensatz einer deutschen Auskunftei verwendet. Dieser enthält Informationen über die Scores der betrachteten Individuen. Daher wollen wir an dieser Stelle näher auf das Scoringssystem der Auskunftei eingehen. Die Auskunftei besitzt für jede gemeldete Person verschiedene Arten von Informationen:

⁷Vgl. z. B. Carey & Hrycay (2001, S. 206) und Henking et al. (2006, Kap. 7.2).

- a) Informationen zur Identifikation, z. B. Name, Adresse, Geburtsdatum,
- b) Informationen zum Zahlungsverhalten, z. B. Anzahl und Art von Kreditaktivitäten, Mobilfunkkonten, Versandhandelskonten,
- c) Informationen über Zahlungsausfälle: Ausfallvariable,
- d) Bonitätsvariablen (Scores) für verschiedene Bereiche, z. B. für Banken, Telekommunikation, Versandhandel.

Die Daten zu den Gruppen a) bis c) erhält die Auskunftsteilnehmer durch ihre Vertragspartner oder aus öffentlichen Schuldnerverzeichnissen. Die Bonitätsvariablen (Scores) berechnet die Auskunftsteilnehmer aus einer Auswahl⁸ von Informationen aus b). Als Grundlage der Berechnung dient eine logistische Regression. Im Folgenden wird die Bestimmung der Scores kurz skizziert.

Bei der logistischen Regression soll eine binäre Variable $Y \in \{0, 1\}$ durch eine gewisse Anzahl von Kovariablen z_k , $k = 1, \dots, K$ erklärt werden.⁹ Dazu wird eine Stichprobe vom Umfang n betrachtet. Das Modell der logistischen Regression unterstellt dann für jeden Merkmalsträger $i = 1, \dots, n$ den Zusammenhang

$$\ln \left(\frac{\pi_i}{1 - \pi_i} \right) = \eta_i$$

oder äquivalent

$$\pi_i = \frac{\exp(\eta_i)}{1 + \exp(\eta_i)} = \frac{1}{\exp(-\eta_i) + 1}$$

zwischen der Erfolgswahrscheinlichkeit $\pi_i = P(Y_i = 1)$ und einer Linearkombination der erklärenden Variablen

$$\eta_i := \beta_0 + \beta_1 z_{i1} + \beta_2 z_{i2} + \dots + \beta_K z_{iK},$$

dem sogenannten linearen Prädiktor. Die Funktionen $g(\pi_i) = \ln \left(\frac{\pi_i}{1 - \pi_i} \right)$ und $h(\eta_i) = \frac{\exp(\eta_i)}{1 + \exp(\eta_i)}$, die die Verknüpfung zwischen π_i und η_i herstellen, werden als Linkfunktio-

⁸Es müssen nicht alle verfügbaren Variablen in die Berechnung einbezogen werden. Einige vorhandene Variablen haben eventuell kaum einen oder gar keinen Einfluss auf das Ausfallverhalten oder sind durch zahlreiche fehlende Werte (*missing values*) gekennzeichnet. Diese Variablen werden dann ebenso nicht berücksichtigt wie Variablen, deren Einfluss auf das Ausfallverhalten anders als theoretisch erwartet und daher nicht plausibel ist. Auch eventuelle Multikollinearität oder Interaktionen zwischen Eingangsvariablen müssen bei der Auswahl berücksichtigt werden. Einleitende Informationen zur Auswahl von Variablen findet man in Henking et al. (2006, Kap. 7.6), ein konkretes Beispiel in Fernandes (2005, S. 9-13).

⁹Die kurze Beschreibung der Logit-Regression orientiert sich an den Monographien von Fahrmeir et al. (2009, Kap. 2.3 und 4.1) und McCullagh & Nelder (1999, Kap. 4). Diese geben darüber hinaus weiterführende Informationen, z. B. zur Schätzung und Interpretation der Regressionskoeffizienten sowie zu alternativen Verfahren, wie der Probit-Regression oder dem komplementären log-log-Modell.

on (im logistischen Regressionsmodell speziell: Logit) bzw. Antwortfunktion (*response function*) bezeichnet. Zur Schätzung der Regressionskoeffizienten β_k , $k = 0, \dots, K$ mittels Maximum-Likelihood-Methode wird für die n Ausfallvariablen Y_i (bedingte) Unabhängigkeit angenommen.

Zu einem gewissen Zeitpunkt in der Vergangenheit wurden die Regressionskoeffizienten β_k , $k = 0, \dots, K$ geschätzt. Dabei diente die binäre Ausfallvariable als Zielvariable und eine Auswahl von Variablen zum Zahlungsverhalten als erklärende Variablen. Die Wahrscheinlichkeit π_i ist im vorliegenden Kontext als Ausfallwahrscheinlichkeit des i -ten betrachteten Individuums zu interpretieren. Die bei der Logit-Regression ermittelten Schätzwerte $\hat{\beta}_k$, $k = 0, \dots, K$ werden benutzt, um – auch zu den folgenden Zeitpunkten – den Rohscore

$$\hat{\eta}_i = \hat{\beta}_0 + \hat{\beta}_1 z_{i1} + \hat{\beta}_2 z_{i2} + \dots + \hat{\beta}_K z_{iK}$$

aus den bekannten individuenspezifischen Variablen z_{ik} , $k = 1, \dots, K$ zu berechnen. Diese Berechnung wird aktualisiert, sobald sich eines der bonitätsrelevanten Merkmale ändert. Aus dem Rohscore kann die individuenspezifische Ausfallwahrscheinlichkeit durch

$$\hat{\pi}_i = h(\hat{\eta}_i) = \frac{\exp(\hat{\eta}_i)}{1 + \exp(\hat{\eta}_i)}$$

geschätzt werden. Der Schätzwert $\hat{\pi}_i$ wird anschließend durch eine monoton fallende Funktion auf den Bereich der natürlichen Zahlen zwischen 1 und 1000 transformiert. Das Ergebnis wird als Score bezeichnet. Bei der Transformation wird die auf den geschätzten Ausfallwahrscheinlichkeiten basierende Risikoordnung umgekehrt: Ein kleiner Score steht für eine hohe geschätzte Ausfallwahrscheinlichkeit und daher niedrige Bonität, ein großer Score für eine niedrige geschätzte Ausfallwahrscheinlichkeit und daher hohe Bonität.

2.1.2 Vergleich von Scoringssystemen

Wie in Abschnitt 2.1.1 angeklungen ist, existiert eine Vielzahl von Scoringmodellen. Innerhalb der verschiedenen Methoden ist dann die Verwendung vieler bonitätserklärender Variablen möglich. Somit stellt sich die Frage, welche Variablen innerhalb welchen Scoringmodells die besten Ergebnisse liefern. Es sind also Kriterien für den Vergleich von Scoringssystemen vonnöten.

Tatsächlich existiert zur Einschätzung von Scoringmethoden eine Vielzahl von Kriterien.¹⁰ Am häufigsten erfolgt die Beurteilung über das Cumulative Accuracy Profile

¹⁰Für einen ersten Überblick vgl. z. B. Krämer & Bücker (2011).

(CAP) oder über die Receiver Operating Characteristic (ROC) und daraus abgeleitete Maße. Daher sollen im Folgenden die Konzepte von CAP und ROC sowie das ROC-gap als abgeleitetes Qualitätskriterium näher beleuchtet werden. Bevor das jedoch geschehen kann, müssen einige Begrifflichkeiten eingeführt werden.

Einleitende Bemerkungen

Wir betrachten eine Menge von n Individuen. Im Folgenden bezeichne die Zufallsvariable X_i den Score des i -ten Individuums. Es wird vereinbart, dass ein großer Score für ein geringes Ausfallrisiko, d. h. für eine gute Bonität, spricht. Die zufällige empirische Verteilungsfunktion der Scores ist¹¹

$$F_n(x) := \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{\{X_i \leq x\}}, \quad x \in \mathbb{R}. \quad (2.1)$$

Sie misst den Anteil der Individuen mit einem Score kleiner oder gleich x unter allen betrachteten Individuen. Die Zufallsvariable Y_i modelliert den Ausfall des i -ten Individuums. Es gelte

$$Y_i = \begin{cases} 1, & \text{wenn das } i\text{-te Individuum ausfällt,} \\ 0, & \text{wenn das } i\text{-te Individuum nicht ausfällt.} \end{cases}$$

Die Zufallsvariablen N_D und N_{ND} zählen die Anzahl der Ausfälle (*defaults*, D) bzw. die Anzahl der Nichtausfälle (*non-defaults*, ND):

$$N_D := \sum_{i=1}^n Y_i \quad \text{und} \quad N_{ND} := n - N_D.$$

Die zufällige **Ausfallquote** (*default rate*)

$$\bar{Y} := \frac{1}{n} \sum_{i=1}^n Y_i = \frac{N_D}{n} \quad (2.2)$$

gibt den Anteil der ausgefallenen Individuen unter allen betrachteten Personen an.

Im Folgenden unterstellen wir, dass es unter den n betrachteten Merkmalsträgern mindestens einen Ausfall und mindestens einen Nichtausfall gibt. Daraus folgt

$$0 < N_D < n, \quad 0 < N_{ND} < n \quad \text{und} \quad 0 < \bar{Y} < 1.$$

¹¹Vgl. z. B. Wasserman (2006, S. 13). Dabei bezeichnet $\mathbb{1}$ die Indikatorfunktion. Sie nimmt den Wert 1 an, falls das interessierende Ereignis, hier $\{X_i \leq x\}$, eintritt, anderenfalls ist sie gleich 0. Zur Indikatorfunktion vgl. z. B. Casella & Berger (2002, S. 113, Definition 3.4.5) oder Karr (1993, S. 19).

Ausfallverhalten		prognostiziert	
		Ausfall $X \leq x$	Nichtausfall $X > x$
beobachtet	Ausfall	Hit	Miss
	Nichtausfall	False Alarm	Correct

Tabelle 2.1: Vergleich von prognostiziertem und beobachtetem Ausfallverhalten

Mit dieser Einschränkung können wir die empirischen Verteilungsfunktionen der Scores der Ausfälle bzw. der Nichtausfälle wie folgt definieren.

- Die **Hit Rate**¹²

$$HR(x) := \frac{1}{N_D} \sum_{i=1}^n \mathbb{1}_{\{X_i \leq x\}} Y_i, \quad x \in \mathbb{R} \quad (2.3)$$

ist die zufällige empirische Verteilungsfunktion der Scores der Ausfälle. Sie misst den Anteil der Ausfälle bis zum Score x unter der Gesamtzahl aller Ausfälle.

- Die **False Alarm Rate**

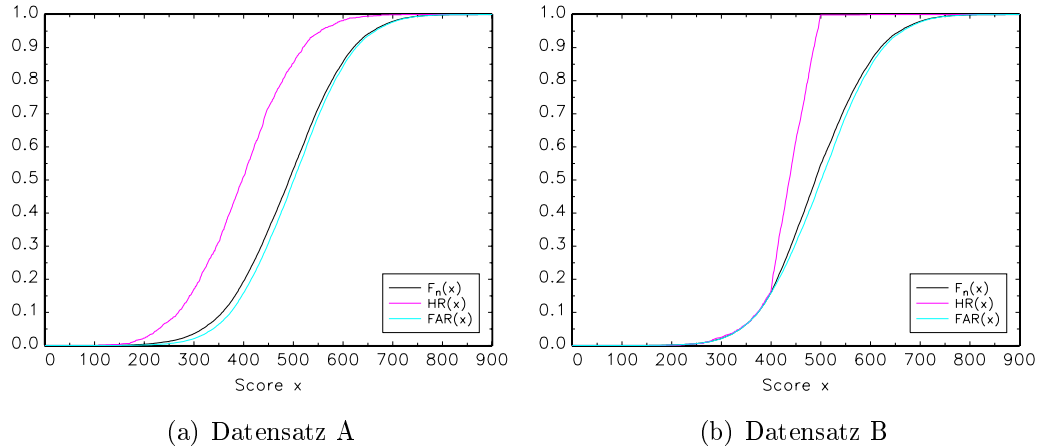
$$FAR(x) := \frac{1}{N_{ND}} \sum_{i=1}^n \mathbb{1}_{\{X_i \leq x\}} (1 - Y_i), \quad x \in \mathbb{R} \quad (2.4)$$

ist die zufällige empirische Verteilungsfunktion der Scores der Nichtausfälle. Sie misst den Anteil der Nichtausfälle bis zum Score x unter der Gesamtzahl aller Nichtausfälle.

Die Bezeichnungen Hit Rate und False Alarm Rate entstammen dem Vergleich von prognostiziertem und tatsächlichem Ausfallverhalten. Dabei erfolgt die Ausfallprognose anhand des sogenannten Cut-off-Scores x . Für Scores, die kleiner oder gleich x sind, wird ein Ausfall prognostiziert. Für Scores, die größer als x sind, wird ein Nichtausfall prognostiziert. Möglich sind vier Fälle, die in Tabelle 2.1 dargestellt sind. Neben den korrekten Entscheidungen (Correct und Hit) gibt es zwei Fehlentscheidungen, die als Fehler 1. Art bzw. Fehler 2. Art bezeichnet werden.¹³ Der Fehler 1. Art liegt vor, wenn ein Nichtausfall prognostiziert, tatsächlich aber ein Ausfall beobachtet wird (Miss). Im umgekehrten Fall (False Alarm) liegt der Fehler 2. Art vor. Die False Alarm Rate $FAR(x)$ beziffert somit die zufällige relative Häufigkeit für den Fehler 2. Art, wenn x

¹²Vgl. hier und im Folgenden Sobehart & Keenan (2001) oder Engelmann et al. (2003).

¹³Vgl. im Folgenden zusätzlich Henking et al. (2006, Abschnitt 7.5, S. 216-229). Dort werden für die Fehlentscheidungen zusätzlich die Begriffe Sensitivität und Spezifität eingeführt, die üblicherweise die (geschätzten) Wahrscheinlichkeiten für die korrekten Entscheidungen bezeichnen. Vgl. hierzu z. B. Hand (1997, S. 132) oder Wehberg et al. (2008, S. 325).



Abbildungung 2.1: Empirische Verteilungsfunktionen in den Beispieldatensätzen

als Cut-off-Score gewählt wird. Die zufällige relative Häufigkeit für den Fehler 1. Art wird durch $1 - HR(x)$ gemessen.¹⁴

Zur Veranschaulichung der vorgestellten Funktionen und auch der folgenden Konzepte dienen zwei Beispieldatensätze:

- Beispieldatensatz A besteht aus insgesamt 15 000 Beobachtungen. Darunter befinden sich 1 500 Ausfälle und 13 500 Nichtausfälle. Daraus ergibt sich eine Ausfallquote von 10 %. Die Scores der Ausfälle sind auf ganze Zahlen gerundete $N(400, 100^2)$ -Zufallszahlen, d. h. Zufallszahlen aus einer Normalverteilung mit Erwartungswert 400 und Standardabweichung 100. Die Scores der Nichtausfälle sind auf ganze Zahlen gerundete $N(500, 100^2)$ -Zufallszahlen.
- Beispieldatensatz B ist genau wie Datensatz A aus 1 500 Ausfällen und 13 500 Nichtausfällen zusammengesetzt. In beiden Gruppen liegen den Scores auf ganze Zahlen gerundete $N(500, 100^2)$ -Zufallszahlen zugrunde. Für die Scores der Ausfälle werden die Zufallszahlen anschließend jedoch so modifiziert,¹⁵ dass sie weitestgehend kleiner als 500 sind.

Abbildungung 2.1 zeigt für beide Datensätze die drei vorgestellten empirischen Verteilungsfunktionen, d. h. die Hit Rate HR , die False Alarm Rate FAR sowie die empirische Verteilungsfunktion aller Scores F_n .

Zwischen den drei empirischen Verteilungsfunktionen besteht natürlich ein Zusammenhang: Die empirische Verteilungsfunktion aller Scores F_n ist eine Konvexkombination der empirischen Verteilungsfunktion der Scores der Ausfälle HR und der empirischen Verteilungsfunktion der Scores der Nichtausfälle FAR . Als Koeffizienten sind die

¹⁴Vgl. auch Fahrmeir et al. (2002, S. 22). Dort werden jedoch nicht wie hier die empirischen Größen, sondern die theoretischen Größen betrachtet. Daher wird dort von Wahrscheinlichkeiten und nicht von Häufigkeiten gesprochen.

¹⁵Von allen Zufallszahlen zwischen 500 und 600 wird nachträglich 100 subtrahiert, von denen zwischen 600 und 700 wird 200 subtrahiert und von denen größer als 700 wird 300 subtrahiert.

Ausfallquote \bar{Y} und die Nichtausfallquote $1 - \bar{Y}$ zu wählen. Formalisiert wird der beschriebene Zusammenhang im folgenden Lemma, welches gleichzeitig die einleitenden Bemerkungen abschließt.

Lemma 2.1 *Sei $0 < \bar{Y} < 1$. Für alle $x \in \mathbb{R}$ gilt dann*

$$F_n(x) = \bar{Y}HR(x) + (1 - \bar{Y})FAR(x). \quad (2.5)$$

Beweis: Wegen $0 < \bar{Y} < 1$ gilt sowohl $N_D > 0$ als auch $N_{ND} > 0$. Daher sind die Funktionen HR und FAR definiert. Für alle $x \in \mathbb{R}$ gilt

$$\begin{aligned} F_n(x) &= \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{\{X_i \leq x\}} \\ &= \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{\{X_i \leq x\}} (Y_i + (1 - Y_i)) \\ &= \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{\{X_i \leq x\}} Y_i + \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{\{X_i \leq x\}} (1 - Y_i) \\ &= \frac{N_D}{n} \frac{1}{N_D} \sum_{i=1}^n \mathbb{1}_{\{X_i \leq x\}} Y_i + \frac{N_{ND}}{n} \frac{1}{N_{ND}} \sum_{i=1}^n \mathbb{1}_{\{X_i \leq x\}} (1 - Y_i) \\ &= \bar{Y}HR(x) + (1 - \bar{Y})FAR(x). \end{aligned}$$

Dabei gilt das letzte Gleichheitszeichen wegen (2.2), (2.3), (2.4) und der Beziehung $\frac{N_{ND}}{n} = \frac{n - N_D}{n} = 1 - \bar{Y}$. \square

CAP – Cumulative Accuracy Profile

Das Cumulative Accuracy Profile (CAP) ist ein Werkzeug zur Einschätzung der Güte eines Scoringsystems. Es wird teilweise auch als Lorenzkurve, Gini curve oder power curve bezeichnet.¹⁶ In der hier betrachteten empirischen Variante ist das CAP definiert als der Polygonzug durch die Punkte

$$(F_n(x), HR(x)), \quad x \in \mathbb{R}.$$

Diese empirische Größe kann als Schätzer für das wahre, unbekannte CAP interpretiert werden. In der Praxis bewegt sich das CAP zwischen zwei Extremfällen:

- Ein triviales Scoringssystem trennt nicht zwischen Ausfällen und Nichtausfällen. Die Verteilungsfunktion aller Scores und die Verteilungsfunktion der Scores der

¹⁶Vgl. Krämer & Bueker (2011).

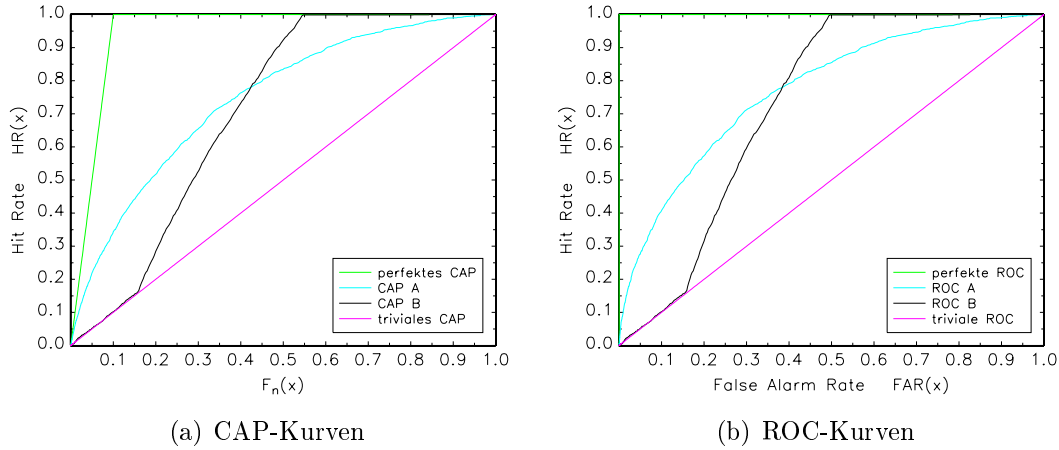


Abbildung 2.2: CAP- und ROC-Kurven der Beispieldatensätze

Ausfälle sind identisch. Folglich verläuft das triviale CAP auf der Winkelhalbierenden zwischen den Punkten $(0, 0)$ und $(1, 1)$. Formal heißt das

$$F_n(x) = HR(x) \text{ für alle } x \in \mathbb{R}.$$

- In einem perfekten Scoringssystem erhalten alle Ausfälle kleinere Scores als der am schlechtesten bewertete Nichtausfall. Folglich besitzt das perfekte CAP zunächst einen linearen Anstieg, bis die Hit Rate den Wert eins erreicht, und verläuft anschließend waagerecht. Formal wird das perfekte CAP durch die folgenden zwei Beziehungen beschrieben:

$$\begin{aligned} 0 \leq F_n(x) < \bar{Y} &\Rightarrow HR(x) = \frac{F_n(x)}{\bar{Y}}, \\ \bar{Y} \leq F_n(x) \leq 1 &\Rightarrow HR(x) = 1. \end{aligned}$$

Je näher das CAP eines Scoringystems am perfekten CAP liegt, desto besser ist das Scoringssystem. In Abbildung 2.2(a) sind die realisierten CAPs der Beispieldatensätze A und B sowie das triviale und das perfekte CAP dargestellt.

ROC – Receiver Operating Characteristic

Ein mit dem CAP eng verwandtes Instrument zur Beurteilung eines Scoringystems ist die Receiver Operating Characteristic (ROC). Sie entsteht aus dem zugehörigen CAP im Grunde durch eine Verlagerung nach links.¹⁷ In der hier betrachteten empirischen Variante ist die ROC definiert als der Polygonzug durch die Punkte

$$(FAR(x), HR(x)), \quad x \in \mathbb{R}.$$

¹⁷Vgl. z. B. Krämer & Bücker (2011, S. 46). Ursächlich ist der Zusammenhang aus Lemma 2.1.

Diese empirische Größe kann als Schätzer für die wahre, unbekannte ROC interpretiert werden.¹⁸ In der Praxis bewegt sich die ROC zwischen zwei Extremfällen:

- Ein triviales Scoringssystem trennt nicht zwischen Ausfällen und Nichtausfällen. Die Verteilungsfunktion der Scores der Ausfälle und die Verteilungsfunktion der Scores der Nichtausfälle sind identisch. Folglich verläuft die triviale ROC auf der Winkelhalbierenden zwischen den Punkten $(0, 0)$ und $(1, 1)$. Formal heißt das

$$FAR(x) = HR(x) \text{ für alle } x \in \mathbb{R}. \quad (2.6)$$

- In einem perfekten Scoringssystem erhalten alle Ausfälle kleinere Scores als der am schlechtesten bewertete Nichtausfall. Folglich ist die Hit Rate gleich eins bevor die False Alarm Rate größer als null wird. In Abbildung 2.2(b), in der die realisierten ROCs der Beispieldatensätze A und B sowie die triviale und die perfekte ROC dargestellt sind, verläuft die perfekte ROC somit auf dem linken und oberen Rand der Grafik. Formal wird die perfekte ROC durch die folgende Beziehung beschrieben:

$$\text{Es existiert ein } x \in \mathbb{R} \text{ mit } HR(x) = 1 \text{ und } FAR(x) = 0. \quad (2.7)$$

Je näher die ROC eines Scoringssystems an der perfekten ROC liegt, desto besser wird das Scoringssystem eingeschätzt.

ROC-gap und Kolmogoroff-Smirnoff-Statistik

Mit Hilfe der ROC und des CAP können jedoch nicht alle Scoringmodelle verglichen werden. Der Grund dafür ist, dass sich ROCs bzw. CAPs schneiden können: Ein Scoringssystem muss „nicht immer gleichmäßig besser oder schlechter als ein anderes Scoreverfahren sein“.¹⁹ Vergleiche dazu auch die ROCs und CAPs der Beispieldatensätze in Abbildung 2.2. Eine vollständige Ordnung der Scoringssysteme kann über skalarwertige Kriterien erreicht werden. Typischerweise werden dabei die Trefferquote (auch *accuracy ratio* oder Gini-Koeffizient) und die AUROC (*area under ROC*) als Maßzahlen verwendet. Da sie im weiteren Verlauf der Arbeit keine Rolle spielen, soll nicht näher auf sie eingegangen werden. Der interessierte Leser sei z. B. auf Engelmann et al. (2003), Henking et al. (2006, Abschnitt 7.5.2, S. 221-229) und Krämer & Bücker (2011) verwiesen. Im Folgenden soll stattdessen das sogenannte ROC-gap vorgestellt werden, da sich zwischen diesem Gütemaß und einer später thematisierten Methode eine Verbindung ergibt.

¹⁸Konfidenzbänder für die ROC werden z. B. in Fahrmeir et al. (2002) behandelt.

¹⁹Fahrmeir et al. (2002, S. 23).

Das ROC-gap²⁰ misst den maximalen vertikalen Abstand zwischen der ROC des betrachteten Systems und der trivialen ROC, d. h.²¹

$$\text{ROC-gap} := \max_{x \in \mathbb{R}} d^{\text{ROC}}(x), \quad (2.8)$$

wobei

$$d^{\text{ROC}}(x) := HR(x) - FAR(x), \quad x \in \mathbb{R}. \quad (2.9)$$

Für die Abstandsfunktion d^{ROC} gilt

$$-1 \leq d^{\text{ROC}}(x) \leq 1, \quad x \in \mathbb{R}.$$

Außerhalb des Beobachtungsbereichs verschwindet der Abstand zwischen Hit Rate und False Alarm Rate, d. h.

$$d^{\text{ROC}}(x) = 0, \text{ falls } x < X_{1:n} \text{ oder } x \geq X_{n:n}.$$

Dabei bezeichnen $X_{1:n}$ und $X_{n:n}$ die kleinste bzw. größte Variable der geordneten Stichprobe $X_{1:n} \leq X_{2:n} \leq \dots \leq X_{n:n}$, also das Minimum bzw. Maximum der Stichprobenvariablen.²² Somit gilt für das ROC-gap

$$0 \leq \text{ROC-gap} \leq 1.$$

Ein triviales Scoringssystem liefert gemäß (2.6) ein ROC-gap von null, ein perfektes Scoringssystem gemäß (2.7) ein ROC-gap von eins. Das dient als Begründung dafür, dass eine Scoringmethode umso besser eingeschätzt wird, je näher das ROC-gap an eins liegt.

Im Folgenden soll der Zusammenhang zwischen dem ROC-gap und dem Kolmogoroff-Smirnoff-Test²³ thematisiert werden. In Krämer & Bücke (2011, S. 50) wird behauptet, dass „der Kolmogoroff-Smirnoff-Test [...] und das ROC-gap [...] monotone Funktionen voneinander“ seien. Das ist ohne Einschränkungen nicht richtig. Blöchliger & Leippold (2006, S. 853) liefern eine mögliche Einschränkung: Falls die ROC konkav ist, d. h. auf ihrem Weg vom Punkt (0, 0) zum Punkt (1, 1) eine Rechtskurve beschreibt,

²⁰Analog zum ROC-gap ließe sich auch ein CAP-gap definieren.

²¹Vgl. Krämer & Bücke (2011, S. 48). In Hoadley & Oliver (1998, S. 56-57) wird nicht der maximale Abstand sondern die Abstandsfunktion, die hier d^{ROC} genannt wird, als ROC-gap bezeichnet.

²²Vgl. z. B. Müller (1991, S. 141) oder Dudewicz & Mishra (1988, S. 280).

²³Gemeint ist der Kolmogoroff-Smirnoff-Homogenitätstest, der teilweise auch Kolmogoroff-Smirnoff-Zweistichprobentest genannt wird. Der Test prüft, ob die Verteilungsfunktionen zweier Zufallsvariablen übereinstimmen. Nicht gemeint ist der Kolmogoroff-Smirnoff-Anpassungstest, der überprüft, ob die unbekannte Verteilungsfunktion einer Grundgesamtheit mit einer vollauffestgelegten stetigen Verteilungsfunktion übereinstimmt.

sind das ROC-gap und die Teststatistik des Kolmogoroff-Smirnoff-Tests (KS-Statistik) zur Überprüfung der Hypothesen

$$H_0 : F^D = F^{ND} \quad \text{versus} \quad H_1 : F^D \neq F^{ND} \quad (2.10)$$

proportional. Hierbei sind F^D und F^{ND} die unbekannten theoretischen Verteilungsfunktionen der Scores der Ausfälle bzw. Nichtausfälle. Je weiter diese Verteilungsfunktionen auseinander liegen, desto besser trennt das Scoringssystem die Ausfälle von den Nichtausfällen. Die Konkavität der ROC ist hinreichend, aber nicht notwendig für die Proportionalität. Des Weiteren ist sie – insbesondere für die empirische Variante der ROC – eine sehr starke Voraussetzung. Eine schwächere hinreichende Bedingung ist die Tatsache, dass die betrachtete ROC niemals unterhalb der trivialen ROC liegt. Sie ist aber ebenfalls nicht notwendig.

Notwendig und hinreichend für die angesprochene Proportionalität zwischen ROC-gap und KS-Statistik ist die Bedingung

$$\max_{x \in \mathbb{R}} d^{\text{ROC}}(x) \stackrel{!}{=} \max_{x \in \mathbb{R}} |d^{\text{ROC}}(x)|, \quad (2.11)$$

denn gemäß (2.8) gilt

$$\text{ROC-gap} = \max_{x \in \mathbb{R}} d^{\text{ROC}}(x)$$

und die KS-Statistik ist ein Vielfaches²⁴ von

$$\max_{x \in \mathbb{R}} |HR(x) - FAR(x)| = \max_{x \in \mathbb{R}} |d^{\text{ROC}}(x)|. \quad (2.12)$$

Bedingung (2.11) gilt genau dann, wenn

$$\max_{x \in \mathbb{R}} d^{\text{ROC}}(x) \geq \max_{x \in \mathbb{R}} -d^{\text{ROC}}(x) = -\min_{x \in \mathbb{R}} d^{\text{ROC}}(x),$$

d. h. wenn der Ausschlag der Abstandsfunktion d^{ROC} in die positiven Zahlen größer ist als der Ausschlag der Funktion in die negativen Zahlen. Wegen dieser Proportionalität zum ROC-gap wird auch die KS-Statistik als Gütemaß für Scoringssysteme herangezogen.²⁵

²⁴ In einem gewissen Teil der Literatur ist die linke Seite von (2.12) bereits die Prüfgröße des Kolmogoroff-Smirnoff-Tests (z. B. Mays (2004a, S. 109), Müller (1991, S. 185-186) oder Rinne (2008, S. 573, Tab. C3/32)). Dann verstärkt sich die Proportionalität zur Identität. In anderen Teilen der Literatur wird der Ausdruck in (2.12) noch mit einem Term, der die jeweiligen Stichprobenumfänge berücksichtigt, multipliziert (z. B. Hartung et al. (2005, S. 520-522)).

²⁵ Vgl. Mays (2004a, S. 109-113) oder Hoadley & Oliver (1998, S. 56-57).

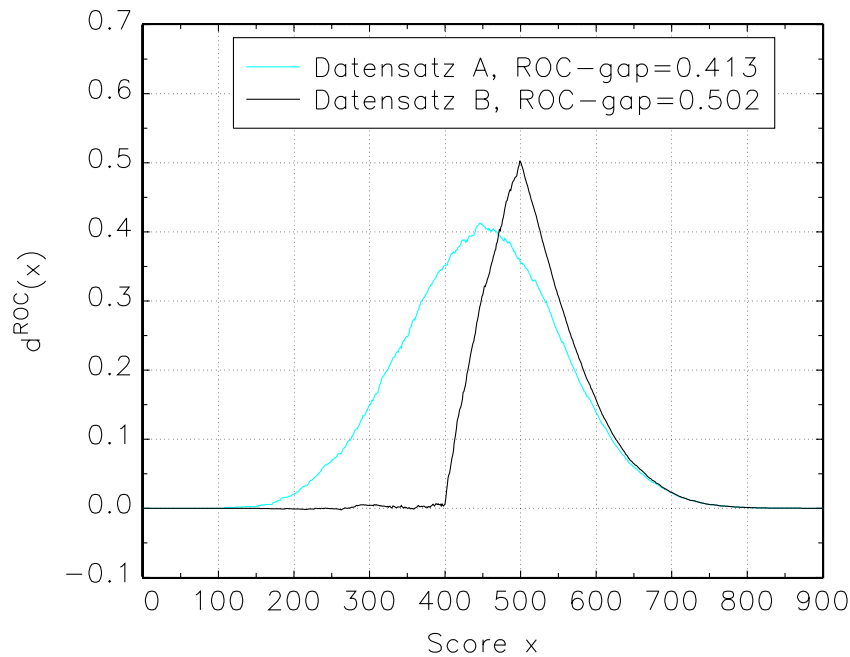


Abbildung 2.3: Empirischer Prozess d^{ROC} für Datensätze A und B

Zum Abschluss dieses Abschnitts vergleichen wir unsere Beispieldatensätze mit Hilfe des ROC-gap. In Abbildung 2.3 ist die Abstandsfunktion d^{ROC} sowohl für Datensatz A als auch für Datensatz B eingezeichnet. Des Weiteren sind die entsprechenden ROC-gaps angegeben. Im Datensatz B ist das ROC-gap bedeutend größer als im Datensatz A, obwohl das Scoringssystem in B bis etwa zum Score 400 keine Trennung von Ausfällen und Nichtausfällen vornimmt²⁶ und daher ungeeignet ist. Als alleiniges Vergleichskriterium sind das ROC-gap und damit auch die KS-Statistik also unzureichend.²⁷ Neben diesem Informationsverlust gegenüber der ROC betrifft ein weiterer Kritikpunkt die Abhängigkeit des ROC-gaps vom zugrunde liegenden Portfolio und vom Zeitpunkt der Auswertung.²⁸

2.2 Ratingklassenbildung

Das zweite Basiselement eines bankinternen Ratingsystems ist die Aufteilung der Kreditnehmer anhand ihrer Scores auf Ratingklassen. Diesem Basiselement ist jedoch im Gegensatz zu den anderen beiden Bausteinen des Ratingsystems in der Literatur bisher wenig Aufmerksamkeit geschenkt worden.²⁹ Dabei basiert die Schätzung der ratingklassenspezifischen Ausfallwahrscheinlichkeit wesentlich auf der Einteilung der Ratingklas-

²⁶Vgl. Abbildung 2.1 und die Bemerkungen zum trivialen Scoringssystem, insbesondere (2.6).

²⁷Vgl. auch Mays (2004a, S. 111-112).

²⁸Vgl. hierzu z. B. Mays (2004b).

²⁹Vgl. auch Krink et al. (2008, S. 2275) oder Foglia et al. (2001, S. 422).

sen. Das dritte Basiselement eines internen Ratingsystems ist also grundlegend vom zweiten abhängig.

Zur Ratingklasseneinteilung gehört zum einen die Wahl einer geeigneten Anzahl von Klassen und zum anderen die Festlegung der Klassengrenzen. Beide Probleme müssen im Hinblick auf das dritte Basiselement in Einklang gebracht werden. So fordern die Basler Richtlinien³⁰

- Homogenität in den Ratingklassen, d. h. Kreditnehmer, die hinsichtlich ihres Risikos ähnlich sind, sollen die gleiche Ratingklasse erhalten,
- Heterogenität zwischen den Ratingklassen, d. h. zwischen den verschiedenen Ratingklassen muss es messbare Unterschiede des Ausfallrisikos geben, sowie
- eine im Hinblick auf die Schätzung der ratingklassenspezifischen Ausfallwahrscheinlichkeiten aussagekräftige Verteilung der Kreditengagements über die Ratingklassen hinweg ohne übermäßige Konzentrationen. Zur Erreichung dieses Ziels werden speziell für Firmen-, Staats- und Bankkredite mindestens sieben Ratingklassen für nicht ausgefallene Kreditnehmer verlangt.

Ferner fordern die Basler Richtlinien, dass die Ratingeinschätzung auf einem Zeithorizont beruhen muss, der ein Jahr übersteigt.³¹

Die Ratingklasseneinteilung bildet die Grundlage für das dritte Basiselement eines bankinternen Ratingsystems. In diesem sind aussagekräftige Schätzwerte für die ratingklassenspezifischen Ausfallwahrscheinlichkeiten zu finden. Solche Schätzwerte sind insbesondere dann nicht zu erwarten, wenn die Ratingklassen nur schwach besetzt sind, d. h. nur wenige Kreditnehmer enthalten. Das ist dann der Fall, wenn es zu viele Ratingklassen gibt oder die Ratingklassen zu schmal sind. Eine zu große Anzahl von Klassen kann ebenfalls dazu führen, dass die Klassen untereinander nicht mehr heterogen sind. Dagegen kommt es bei zu breiten Ratingklassen, also insbesondere dann, wenn die Klassenanzahl zu gering gewählt ist, zu Inhomogenität in den Klassen. Wird dann für alle Kreditnehmer einer Ratingklasse eine einheitliche Ausfallwahrscheinlichkeit verwendet, kann es zu einer starken Über- oder Unterschätzung des individuellen Risikos kommen. Dies führt zur sogenannten Negativauslese (*adverse selection*), d. h. nur die relativ schlechtesten Kunden der Ratingklasse verbleiben beim Kreditinstitut, weil sie vergleichsweise günstige Konditionen erhalten. Die relativ besten Kunden erhalten vergleichsweise schlechte Konditionen und wandern zu anderen Banken ab.³²

³⁰Vgl. BCBS (2006, S. 90-92, Paragraphen 397-410).

³¹Vgl. BCBS (2006, S. 93, Paragraph 414).

³²Vgl. Pfingsten & Schröck (2000, S. 18-19) und Krink et al. (2008, S. 2276).

2.2.1 Methoden

Bevor im dritten Kapitel Methoden der Bruchpunktschätzung vorgestellt werden, die die Scores zu optimalen Ratingklassen zusammenfassen, werden an dieser Stelle kurz vorhandene Ansätze beschrieben. Dabei wird unterschieden, ob die Ausfallwahrscheinlichkeiten der Ratingklassen vorgegeben sind oder nicht. Allen Ideen ist gemein, dass die Anzahl der Ratingklassen vorgegeben sein muss.

Vorgegebene Risikoniveaus

In manchen Fällen existieren für die Ausfallwahrscheinlichkeiten der Ratingklassen Referenzgrößen, die für das eigene Ratingsystem übernommen werden sollen. Sie stammen z. B. von einer dem Kreditinstitut übergeordneten Institution oder von Ratingagenturen. Eine Voraussetzung für die sinnvolle Verwendung der Referenzgrößen ist es, dass die Datenstruktur von Referenzportfolio und eigenem Portfolio vergleichbar ist.³³ Dies betrifft unter anderem

- die Definition des Ausfallereignisses (Zahlungsverzögerungen, Zahlungsunfähigkeit),
- den Typ der Engagements (Konsument, Unternehmen, Land),
- deren Größe (Forderungshöhen) und
- deren Zusammensetzung (Branchen, Regionen).

Beispielhaft für die Übernahme der Ratingklassen und Ausfallwahrscheinlichkeiten einer Dachorganisation seien Sparkassen genannt, die die Masterskala des Deutschen Sparkassen- und Giroverbands (DSGV) verwenden.³⁴ Sollen die Ratingklassen von Ratingagenturen als Bezugspunkt dienen, werden die regelmäßig veröffentlichten langfristigen Ausfallquoten³⁵ als Risikoniveaus der eigenen Ratingklassen verwendet. Kritisch wird bei diesem sogenannten *external mapping* neben den oben genannten Punkten außerdem gesehen, dass Ratingagenturen im Gegensatz zu Banken keine Point-in-Time-Ratings vergeben. Die Ratingagenturen bewerten vielmehr über den gesamten Konjunkturzyklus hinweg (Through the Cycle, TTC). Teilweise wird auch geäußert, dass die TTC-Bewertungen zusätzlich auf Stressszenarien beruhen (Stress-TTC).³⁶ Ferner sind folgende Punkte zu beachten:

³³Vgl. z. B. Fernandes (2005, S. 39).

³⁴Vgl. Fischer (2002, S. 242, Abb. 2) für die Ausfallwahrscheinlichkeiten der Ratingklassen des DSGV.

³⁵Siehe z. B. Moody's Investors Service (2012), im Folgenden kurz: Moody's (2012).

³⁶Vgl. z. B. Krink et al. (2008, S. 2277) oder Höse (2007, S. 13).

- Von den Ratingagenturen werden durchschnittliche Ausfallquoten über verschiedene Zeiträume berechnet und veröffentlicht. Bei Moody's (2012) existieren beispielsweise Ausfallquoten für die Zeiträume 1920-2011 und 1970-2011. Welcher Zeitraum gilt als Referenz?
- Die Ausfallquoten können sich mit jeder neuen Veröffentlichung der Ratingagentur ändern. Wird das eigene Ratingsystem deswegen jährlich angepasst?
- Stimmen die der Ausfallwahrscheinlichkeit bzw. -quote zugrunde liegenden Zeithorizonte überein? Bei Moody's (2012) existieren beispielsweise Einjahres- und Fünfjahresausfallquoten.

Sind die Referenzwahrscheinlichkeiten dann festgelegt, gilt es, die eigenen Kreditnehmer einer dieser Größen und damit einer Ratingklasse zuzuordnen. Dafür existieren verschiedene mehr oder weniger fortgeschrittene Ansätze.

Der folgende naive Ansatz setzt voraus, dass für alle Individuen ein individueller Schätzwert für die Ausfallwahrscheinlichkeit, z. B. aus der Logit-Regression, vorliegt. Dieser sei mit $\hat{\pi}_i$ bezeichnet. Ein Individuum wird nun der Referenzwahrscheinlichkeit zugeordnet, der der eigene Schätzwert $\hat{\pi}_i$ am nächsten liegt. Als Trennstellen zwischen den Ratingklassen werden also die Mittelpunkte der benachbarten Risikoniveaus der Masterskala verwendet.³⁷ Dies sei beispielhaft an den DSGV-Ratingklassen 13, 14 und 15 mit den zugehörigen Ausfallwahrscheinlichkeiten 10 %, 15 % bzw. 20 % verdeutlicht.³⁸ Alle Individuen mit einer geschätzten Ausfallwahrscheinlichkeit zwischen 12.5 % und 17.5 % würden hier der Ratingklasse 14 zugeschlagen. Somit sind 12.5 % und 17.5 % gerade die Trennstellen zur Ratingklasse 13 bzw. 15.

Die beiden folgenden Optimierungsansätze zur Bestimmung der Trennstellen benötigen keine individuellen PD-Schätzwerte. Fernandes (2005, S. 39) schlägt vor, die Summe der quadrierten Abweichungen zwischen

- den vorgegebenen Risikoniveaus der K Ratingklassen und
- den realisierten Ausfallquoten der K Klassen des eigenen Modells

zu minimieren. Die vorgegebenen Risikoniveaus der K Ratingklassen werden mit y_k^B bezeichnet, wobei $k = 1, \dots, K$ gelte und das B auf die Referenzskala (*benchmark*) hinweise. Die konkreten Ausfallquoten der K Klassen des eigenen Modells berechnen sich gemäß

$$\bar{y}_k = \frac{1}{n_k} \sum_{i=1}^{n_k} y_i, \quad k = 1, \dots, K,$$

³⁷Vgl. z. B. Krink et al. (2008, S. 2277).

³⁸Vgl. Fischer (2002, S. 242, Abb. 2).

wobei y_i die Realisation der i -ten Ausfallvariable Y_i und n_k die Anzahl der Kreditnehmer in der k -ten Ratingklasse bezeichnet. Formal ausgedrückt lautet der Ansatz von Fernandes also

$$\sum_{k=1}^K (y_k^B - \bar{y}_k)^2 \rightarrow \min_{n_1, \dots, n_{K-1}} ! \quad (2.13)$$

In der Arbeit von Fernandes wird dabei über die Besetzungszahlen n_k der einzelnen Ratingklassen minimiert. Diese sind in (2.13) implizit in den Ausfallquoten \bar{y}_k enthalten. Möglich ist stattdessen aber auch die Verwendung der $K-1$ Trennstellen zwischen den Klassen. Diese Trennstellen legen die Besetzungszahlen n_k eindeutig fest.

Anderson (2007, S. 422) modifiziert für seinen *benchmark breakpoints* genannten Ansatz die Methode von Fernandes. Er verwendet statt der Ausfallwahrscheinlichkeiten und -quoten die entsprechenden logarithmierten Chancen (*log odds*). Zusätzlich wird die jeweilige Klassenbesetzung berücksichtigt. Formal lautet der Ansatz:³⁹

$$\sum_{k=1}^K n_k \left(\ln \left(\frac{1 - y_k^B}{y_k^B} \right) - \ln \left(\frac{1 - \bar{y}_k}{\bar{y}_k} \right) \right)^2 \rightarrow \min_{s_1, \dots, s_{K-1}} ! \quad (2.14)$$

Dabei bezeichnen s_1, \dots, s_{K-1} die Trennstellen zwischen den K Klassen (*score breaks*). Diese Trennstellen legen die Besetzungszahlen n_k fest und sind damit implizit auch in den Ausfallquoten \bar{y}_k enthalten. Die weitere Notation bleibt unverändert.

Bemerkung 2.2 (Logarithmierte Chancen im Anderson-Ansatz)

Für alle $p, q \in]0, 1[$ gilt

$$\left(\ln \left(\frac{1-p}{p} \right) - \ln \left(\frac{1-q}{q} \right) \right)^2 = \left((-1) \cdot \left(\ln \left(\frac{1-p}{p} \right) - \ln \left(\frac{1-q}{q} \right) \right) \right)^2 \quad (2.15)$$

$$= \left(\ln \left(\frac{p}{1-p} \right) - \ln \left(\frac{q}{1-q} \right) \right)^2 \quad (2.16)$$

$$= \left(\ln \left(\frac{\frac{p}{1-p}}{\frac{q}{1-q}} \right) \right)^2. \quad (2.17)$$

Die Struktur auf der linken Seite von (2.15) ist identisch mit der Struktur in (2.14). Durch die Umformung in (2.16) ist gezeigt, dass tatsächlich von einer Differenz von logarithmierten Chancen (*log odds* bzw. *Logits*) gesprochen werden kann. Erst jetzt ist nämlich die üblicherweise verwendete Form der Chance (*odds*) sichtbar. Mit (2.17)

³⁹In der zitierten Gleichung aus Anderson (2007, S. 422, Equation 20.2) befindet sich ein Schreibfehler: Statt s_{k-1} muss es korrekterweise s_{q-1} heißen. Ferner muss es in der Zeile unter der Formel nicht „odds“ sondern „default rates“ heißen. Letzteres ergibt sich aus der Beschreibung der Formel auf den Zeilen 6 bis 9.

wird deutlich, dass das Kriterium von Anderson auf den sogenannten relativen Chancen (odds ratio oder auch Kreuzproduktverhältnis) beruht.⁴⁰

Bemerkung 2.3 (Vergleich der Ansätze von Anderson und Fernandes)

Anderson (2007, S. 422) behauptet, dass seine Entscheidungsregel (2.14) besser geeignet sei als die von Fernandes, „if there is an exponential increase in risk between grades“. Er liefert dafür aber weder Motivation noch Beleg noch Beweis. Unklar ist auch, in welchem Sinne seine Methode besser sein soll. Offensichtlich liefert sie die beste Trennstelle im Sinne von (2.14), wohl kaum aber im Sinne von (2.13).

Bemerkung 2.4 (Einfluss der Klassenbesetzung)

Im Ansatz von Anderson wird im Gegensatz zum Ansatz von Fernandes die Besetzung der Klassen n_k berücksichtigt. Untersuchenswert wäre die Frage, inwiefern das Weglassen bzw. die Hinzunahme des Faktors n_k die Ergebnisse der Minimierungsaufgaben beeinflusst. Antworten geben die Ergebnisse von Abschnitt 4.2.1, in dem auch die entsprechend modifizierten Methoden betrachtet werden.

Unbekannte Risikoniveaus

Sind keine Referenzwerte für die Ausfallwahrscheinlichkeiten der Ratingklassen vorgegeben, müssen andere Ansätze gefunden werden, um die Trennstellen zwischen den Ratingklassen festzulegen. Die einfachste Variante teilt den Scorebereich in gleich große Stücke. Bei einem diskreten Scorebereich von 1 bis 1000 und 10 Ratingklassen ergäben sich $[1, 100]$, $[101, 200]$, \dots , $[901, 1000]$ als Ratingklassen. Wenn jedoch der Score mit der geschätzten individuellen Ausfallwahrscheinlichkeit übereinstimmt, ist dieses Verfahren viel zu einfach gestrickt. Wird nämlich das Intervall $[0, 1]$ in z. B. 10 Ratingklassen geteilt, ergeben sich die Trennstellen 0.1, 0.2, \dots , 0.8 und 0.9. Folglich würde sich ein großer Anteil der Individuen in den besten Klassen mit geringen Ausfallwahrscheinlichkeiten konzentrieren. Die schlechtesten Klassen mit hohen Ausfallwahrscheinlichkeiten wären dagegen dünn besetzt.⁴¹

Eine weitere sehr einfache Methode der Ratingklasseneinteilung besteht darin, die Individuen gleichmäßig auf die Ratingklassen zu verteilen.⁴² Die entsprechende Regel lautet: Verwende für K Ratingklassen die $\frac{1}{K}$ -Quantile der empirischen Scoreverteilung als Trennstellen zwischen den Ratingklassen. Bei z. B. 10 Ratingklassen wären dann die Dezile $\tilde{x}_{0.1}$, $\tilde{x}_{0.2}$, \dots , $\tilde{x}_{0.9}$ die Klassengrenzen. Anstelle eine gleichmäßige Verteilung der Individuen auf die Ratingklassen anzustreben, besteht auch die Möglichkeit, sich eine andere Verteilung über die K Ratingklassen vorzugeben. Die Vermutung liegt

⁴⁰Zur Definition der genannten Begriffe vgl. z. B. Toutenburg (2003, S. 387,396), Henking et al. (2006, S. 294) oder Tutz (2000, S. 37,47).

⁴¹Vgl. auch Krink et al. (2008, S. 2277, Fußnote 6).

⁴²Vgl. Krink et al. (2008, S. 2276-2277).

nahe, dass die Ratingklassen zum Schufa-Branchenscore für Sparkassen auf diese Weise entstanden sind. Dort enthalten⁴³

- die besten zwei Klassen einen Anteil von ca. 20 % der Population,
- die folgenden fünf Klassen einen Anteil von ca. 10 % der Population,
- die folgende Klasse einen Anteil von ca. 5 % der Population,
- die folgende Klasse einen Anteil von ca. 2 % der Population,
- die schlechtesten drei Klassen einen Anteil von ca. 1 % der Population.

Die Trennstellen zwischen den Ratingklassen sind also die Quantile der Scoreverteilung zu den Niveaus 1 %, 2 %, 3 %, 5 %, 10 %, 20 %, 30 %, 40 %, 50 %, 60 % und 80 %.

Als erster fortgeschrittener Ansatz zur Festlegung der Ratingklassenstruktur sei das K-Means-Verfahren der Clusteranalyse genannt. Das Grundprinzip der Clusteranalyse ist die Homogenität in den Klassen und Heterogenität zwischen den Klassen.⁴⁴ Es stimmt mit den Anforderungen des Basler Ausschusses an ein Ratingsystem überein. Demzufolge erscheint die Clusteranalyse als ein geeigneter Kandidat zur Einteilung der Ratingklassen. Als spezielles Verfahren innerhalb des Gebiets der Clusteranalyse wird der K-Means-Algorithmus vorgeschlagen und verwendet.⁴⁵ Der von einem Startwert ausgehende, iterative Algorithmus wird auf den geschätzten individuellen Ausfallwahrscheinlichkeiten $\hat{\pi}_i$, die sich durch den Scoringprozess ergeben, angewendet. Das Ziel des Verfahrens ist es, eine Zerlegung der Kreditnehmer zu finden, die die Streuung innerhalb der K Cluster, d. h. hier der Ratingklassen, minimiert:

$$\sum_{k=1}^K \sum_{i=1}^{n_k} (\hat{\pi}_i - \bar{\pi}_k)^2 \rightarrow \min!$$

Hierbei bezeichnet n_k wiederum die Anzahl der Individuen in der k -ten Ratingklasse und

$$\bar{\pi}_k = \frac{1}{n_k} \sum_{i=1}^{n_k} \hat{\pi}_i$$

beschreibt als Clusterzentrum die mittlere geschätzte Ausfallwahrscheinlichkeit in Ratingklasse k . Von Vorteil ist das K-Means-Verfahren, weil es einfach zu implementieren und schnell ist. Größter Nachteil ist die Konvergenz gegen lokale Optima oder anders

⁴³Vgl. Korczak & Wilken (2009, S. 8-9). Zur Qualität der in dieser Studie gezogenen Schlussfolgerungen vgl. Huschens (2010).

⁴⁴Vgl. Bacher et al. (2010, S. 16) oder Everitt et al. (2011, S. 7).

⁴⁵Vgl. z. B. Foglia et al. (2001) und Fernandes (2005). Dort wird das K-Means-Verfahren auch jeweils kurz skizziert. Eine ausführliche Behandlung findet man z. B. in Bacher et al. (2010, Kap. 12).

ausgedrückt die Abhängigkeit vom gewählten Startwert.⁴⁶ Ferner ist zu bemerken, dass die Clusterzentren das primäre Interpretationsziel des K-Means-Verfahrens sind.⁴⁷ In der hier vorgestellten Anwendung sind aber vielmehr die Klassengrenzen von Interesse.

Krink et al. (2008, S. 2278) liefern zwei weitere Ansätze zur Festlegung der Ratingklassen. Für beide werden individuelle Schätzungen für die Ausfallwahrscheinlichkeit (PD) benötigt. Die Grundlage für die Verfahren bildet die Feststellung, dass sich bei der Ratingklassenbildung Unterschiede zwischen den geschätzten individuellen PDs und der geschätzten PD der Ratingklasse ergeben.

Der erste beschriebene Ansatz minimiert die Summe der quadrierten Abweichungen zwischen den geschätzten individuellen Ausfallwahrscheinlichkeiten $\hat{\pi}_i$ und der geschätzten PD der zugehörigen Ratingklasse \overline{PD}_k . Er wird formal wie folgt beschrieben:⁴⁸

$$\sum_{k=1}^K \sum_{i=1}^{n_k} (\overline{PD}_k - \hat{\pi}_i)^2 \rightarrow \min! \quad (2.18)$$

Die Zahl n_k bezeichnet wiederum die Anzahl der Individuen in der k -ten Ratingklasse.

Bemerkung 2.5 (Zu den geschätzten Ausfallwahrscheinlichkeiten \overline{PD}_k)

Wie \overline{PD}_k berechnet wird, wird bei der Lektüre von Krink et al. nicht vollends klar. Die Beschreibung „ \overline{PD}_b is the pooled PD for risk bucket b “ schließt keine der beiden folgenden, sinnvollen Berechnungsvorschriften aus:⁴⁹

- (i) \overline{PD}_k ist das arithmetische Mittel der individuellen PD-Schätzwerte in Ratingklasse k , d. h.

$$\overline{PD}_k = \bar{\pi}_k = \frac{1}{n_k} \sum_{i=1}^{n_k} \hat{\pi}_i,$$

- (ii) \overline{PD}_k ist die Ausfallquote der Ratingklasse k , d. h.

$$\overline{PD}_k = \bar{y}_k = \frac{1}{n_k} \sum_{i=1}^{n_k} y_i.$$

⁴⁶Vgl. Krink et al. (2008, S. 2277) und Bacher et al. (2010, S. 301). Die mehrmalige Durchführung des Verfahrens mit verschiedenen zufälligen Startwerten wird als Lösungsansatz aufgeführt. Vgl. dazu Bacher et al. (2010, S. 305) oder Everitt et al. (2011, S. 125-126).

⁴⁷Vgl. Bacher et al. (2010, S. 314).

⁴⁸Vgl. Krink et al. (2008, S. 2278, Formel (1)).

⁴⁹Auch die Wortgruppe „the objective function [...] is similar to the one underlying the k-means algorithm“ (Krink et al., 2008, S. 2278) liefert keinen weiteren Hinweis, da *similar* sowohl mit ähnlich als auch mit gleich übersetzt werden kann. Zur Motivation der zweiten Berechnungsvorschrift vgl. auch Abschnitt 2.3.

Im erstgenannten Fall stimmt die Zielfunktion in (2.18) mit der Clusteranalyse überein. Die Algorithmen zur Minimierung sind jedoch unterschiedlich.⁵⁰

Der zweite Ansatz von Krink et al. (2008, S. 2278, Formel (6)) adressiert das Adverse-Selection-Phänomen, indem anstelle der geschätzten Ausfallwahrscheinlichkeiten die Zinssätze, die sich aus den PD-Schätzwerten ergeben, verwendet werden. Minimiert wird alsdann die Summe der quadrierten Preisabweichungen, d. h.

$$\sum_{k=1}^K \sum_{i=1}^{n_k} (r(\overline{\text{PD}}_k) - r(\hat{\pi}_i))^2 \rightarrow \min! \quad (2.19)$$

Dabei bezeichnet $r(\pi)$ den Zinssatz, der sich bei einer Ausfallwahrscheinlichkeit von π ergibt. Details zur Berechnungsvorschrift für r werden an dieser Stelle ausgelassen. Sie können im zitierten Papier nachgelesen werden. Das in Bemerkung 2.5 geschilderte Problem besteht bei diesem Ansatz weiterhin.

Bemerkung 2.6 (Erweiterung der Ansätze und Lösungsalgorithmus)

- (i) In den beiden zuletzt beschriebenen Ansätzen gehen alle Individuen mit demselben Gewicht in die Minimierung ein. Um die jeweiligen Forderungshöhen zu berücksichtigen, können diese als Gewichte in den einzelnen Summanden ergänzt werden. Vergleiche dazu die Formeln (1b) und (6b) in Krink et al. (2008).
- (ii) Krink et al. (2008) lösen die genannten Optimierungsprobleme mithilfe der sogenannten Differential Evolution⁵¹. Ein besonderer Vorteil dieses Algorithmus – und damit ein bemerkenswerter Unterschied zum K-Means-Verfahren der Clusteranalyse – ist die unproblematische Behandlung von Nebenbedingungen. So wird im zitierten Papier den Anforderungen der Basler Richtlinien Rechnung getragen, indem⁵²
 - (a) als Untergrenze für das Ausfallrisiko einer Ratingklasse der Wert 0.03% vereinbart sowie
 - (b) eine minimale und maximale Klassenbesetzung festgelegt wird.

2.2.2 Vergleichskriterien

Zum Vergleich und zur Beurteilung der verschiedenen Ansätze existieren verschiedene Kriterien. So wird im zuletzt zitierten Papier von Krink et al. (2008, S. 2280, Table 4) die Summe der quadrierten Bepreisungsfehler, d. h. die Doppelsumme in (2.19) und damit das Optimalitätskriterium, als Qualitätsindikator verwendet. Insofern ist es nicht

⁵⁰Vgl. hierzu Bemerkung 2.6(ii).

⁵¹Näheres zu dieser Methode kann der interessierte Leser in Storn & Price (1997) oder Price et al. (2005) finden. Einen kurzen Überblick gibt Appendix A in Krink et al. (2008, S. 2284-2285).

⁵²Vgl. Krink et al. (2008, S. 2278-2279).

verwunderlich, dass der Minimierungsansatz (2.19) in der Studie am besten abschneidet. Problematisch an diesem Vergleichskriterium – insbesondere im Hinblick auf die später in dieser Arbeit untersuchte Datenmenge – ist, dass dafür die geschätzten individuellen Ausfallwahrscheinlichkeiten benötigt werden.

Dies ist bei der im Folgenden vorgestellten Calinski-Harabasz-Statistik (CH-Statistik, auch: *Variance Ratio Criterion*, VRC) nicht der Fall. Diese ursprünglich aus der Clusteranalyse stammende Methode zur Bestimmung einer optimalen Anzahl von Clustern wurde zur Bestimmung einer optimalen Anzahl von Ratingklassen auch in das Kreditrisiko übernommen.⁵³ In beiden Fällen wird dafür die CH-Statistik über die Clusteranzahl K maximiert. Die Güte des Verfahrens innerhalb der Clusteranalyse wurde in der häufig zitierten Arbeit von Milligan & Cooper (1985) bestätigt. Dort schnitt die CH-Statistik unter 30 Kriterien zur Bestimmung der Clusterzahl am besten ab. Laut Anderson (2007, S. 421) kann die CH-Statistik aber auch zum Vergleich konkurrierender Ratingoptionen eingesetzt werden.

Die CH-Statistik wird formal für mehrdimensionale Beobachtungen aufgestellt. Bei Anderson werden jedoch nur univariate Daten, nämlich die Ausfallindikatoren und daraus abgeleitete Ausfallquoten, verwendet. In diesem Fall vereinfacht sich die CH-Statistik zu

$$CH(K) = \frac{\frac{1}{K-1} \sum_{k=1}^K n_k (\bar{Y}_k - \bar{Y})^2}{\frac{1}{n-K} \sum_{k=1}^K \sum_{j=1}^{n_k} (Y_{kj} - \bar{Y}_k)^2}, \quad (2.20)$$

wobei

- n die Anzahl der Beobachtungen,
- K die Anzahl der Klassen,
- n_k die Anzahl der Beobachtungen in Klasse k (=Besetzung der k -ten Klasse),
- Y_{kj} den Ausfallindikator der j -ten Beobachtung in der k -ten Klasse,
- $\bar{Y}_k = \frac{1}{n_k} \sum_{j=1}^{n_k} Y_{kj}$ die Ausfallquote der k -ten Klasse und
- $\bar{Y} = \frac{1}{n} \sum_{k=1}^K \sum_{j=1}^{n_k} Y_{kj} = \frac{1}{n} \sum_{i=1}^n Y_i$ die Gesamtausfallquote

bezeichnet.

⁵³Vgl. Anderson (2007, S. 420) und Fernandes (2005, S. 35).

Bemerkung 2.7 (Literatur zur CH-Statistik und Fehler)

Die CH-Statistik beruht auf der Arbeit von Calinski & Harabasz (1974). Neuere und gut lesbare Darstellungen sind z. B. in Maulik & Bandyopadhyay (2002, S. 1651), Vendramin et al. (2009, S. 735), Everitt et al. (2011, S. 127 in Verbindung mit S. 114)⁵⁴ oder Bacher et al. (2010, S. 307 in Verbindung mit S. 299) zu finden. Beachte dabei, dass $\|\cdot\|^2$ und $(\cdot)(\cdot)^T$ im eindimensionalen Fall zu $(\cdot)^2$ werden sowie dass die Spur⁵⁵ (trace) im eindimensionalen Fall gleich der (1×1) -Matrix, also der Skalar selbst ist. In den Arbeiten von Anderson (2007, S. 420, Gleichung 20.1) und Fernandes (2005, S. 35) befindet sich jeweils ein Fehler im Nenner von (2.20). Dort wird die quadrierte Differenz jeweils fälschlicherweise mit der Gruppengröße (n_k bzw. n_i) multipliziert. Durch diesen Fehler scheint es fraglich, ob die CH-Statistik, wie von Anderson angegeben, wirklich zum Vergleich unterschiedlicher Ratingoptionen geeignet ist.

Inhaltlich setzt die CH-Statistik in (2.20) die Interklassenstreuung im Zähler und die Innerklassenstreuung im Nenner ins Verhältnis. Große Werte zeugen also von in sich homogenen und untereinander heterogenen Klassen und sind demnach wünschenswert. Da im hier betrachteten Spezialfall alle Beobachtungswerte entweder gleich null oder gleich eins sind, d. h. $Y_{kj} \in \{0, 1\}$ gilt, ergibt sich für den Nenner eine alternative Darstellung. Es gilt

$$\frac{1}{n - K} \sum_{k=1}^K \sum_{j=1}^{n_k} (Y_{kj} - \bar{Y}_k)^2 = \frac{1}{n - K} \sum_{k=1}^K n_k \bar{Y}_k (1 - \bar{Y}_k), \quad (2.21)$$

denn wegen $Y_{kj}^2 = Y_{kj}$ und der Verschiebungsdarstellung ist⁵⁶

$$\frac{1}{n_k} \sum_{j=1}^{n_k} (Y_{kj} - \bar{Y}_k)^2 = \frac{1}{n_k} \sum_{j=1}^{n_k} Y_{kj}^2 - \bar{Y}_k^2 = \frac{1}{n_k} \sum_{j=1}^{n_k} Y_{kj} - \bar{Y}_k^2 = \bar{Y}_k - \bar{Y}_k^2 = \bar{Y}_k (1 - \bar{Y}_k).$$

Im vorliegenden eindimensionalen Fall ist die rechte Seite von Gleichung (2.20) auch als Teststatistik der univariaten Varianzanalyse (ANOVA) bekannt, die unter bestimmten Voraussetzungen, die hier nicht erfüllt sind, F -verteilt ist.⁵⁷ Daher wird die Statistik CH teilweise auch als F -Wert bezeichnet.⁵⁸

Bemerkenswert ist, dass auch die CH-Statistik ein Kriterium zur Bestimmung der Trennstellen zwischen Ratingklassen sein kann. Wird die Klassenanzahl fest vorgege-

⁵⁴In der dritten Auflage (Everitt, 1993) ist die Statistik CH zwar formal korrekt definiert (S. 101, Gleichung (5.13)), jedoch taucht in der Formel für \mathbf{W} auf S. 91, Gleichung (5.1), bereits der Faktor $\frac{1}{n-g}$ auf, der dort nicht hingehört. Analoges gilt für die Formel für \mathbf{T} . Die Gleichung für \mathbf{B} ist korrekt.

⁵⁵Die Spur einer $(n \times n)$ -Matrix ist gleich der Summe ihrer n Hauptdiagonalelemente.

⁵⁶Vgl. Mosler & Schmid (2011, S. 183).

⁵⁷Vgl. z. B. Toutenburg (2003, Kap. 5, S. 195-205), Hartung et al. (2005, Kap. XI, S. 609-612) oder Rice (2007, Kap. 12.2, S. 477-485). Die nicht erfüllte Voraussetzung ist die i.i.d.-Normalverteilung der Fehlervariablen und – daraus folgend – die Normalverteilung der Beobachtungen in den jeweiligen Gruppen.

⁵⁸Vgl. Bacher et al. (2010, S. 307).

ben, führt die Maximierung der CH-Statistik über die Trennstellen zu einer in diesem Sinne optimalen Ratingklasseneinteilung. Genauso wie die in Abschnitt 2.2.1 beschriebenen Optimierungsansätze hat aber auch diese Idee ein Problem: Die Festlegung der Klassengrenzen erfolgt zwar anhand eines in gewisser Weise sinnvollen, in gewisser Weise jedoch auch willkürlich gewählten Optimierungskriteriums. Es existiert kein Modell, auf dessen Basis entschieden werden könnte. Das ist bei den in Kapitel 3 vorgestellten Methoden anders. Doch bevor der Fokus dahin geht, erfolgt noch die Vorstellung des dritten Basiselements des Ratingsystems.

2.3 Risikoquantifizierung auf Ratingklassenebene

Der dritte grundlegende Baustein eines bankinternen Ratingsystems ist die Risikoquantifizierung auf der Ebene der Ratingklassen. Hier verlangt der Basler Ausschuss für Bankenaufsicht in jedem Fall die Schätzung der Ausfallwahrscheinlichkeiten in den einzelnen Klassen.⁵⁹ Die resultierenden Schätzwerte dienen später bei der Berechnung der Eigenkapitalunterlegung als einheitliche Ausfallwahrscheinlichkeit aller Kreditnehmer der jeweiligen Ratingklasse.⁶⁰ Weil dieses Element des Ratingsystems in der Literatur bereits vielfältig besprochen wurde und im Folgenden eine untergeordnete Rolle spielt, bleiben die Ausführungen an dieser Stelle eher kurz.

Der kanonische Schätzer für die Ausfallwahrscheinlichkeit ist die zufällige Ausfallquote. Im Falle stochastisch unabhängiger Ausfallereignisse hat die Ausfallquote viele wünschenswerte Eigenschaften, z. B. die Konsistenz gegen die Ausfallwahrscheinlichkeit. Die Unabhängigkeitsannahme ist aber eine trivialisierende Annahme, die im Retailgeschäft eventuell noch gerechtfertigt, bei Krediten an Unternehmen oder Staaten jedoch unzulässig ist. Hier besitzen die Ausfallereignisse im Regelfall eine positive Korrelation. Dadurch können der Ausfallquote zahlreiche wertvolle Eigenschaften wie die Konsistenz verloren gehen. Insofern ist die Verwendung reiner Querschnittsdaten aus einer Periode ungeeignet zur Schätzung der langfristigen Ausfallwahrscheinlichkeit einer Ratingklasse. Eine bessere Alternative ist daher die Verwendung von Längsschnittdaten, d. h. von Beobachtungen aus mehreren Perioden.⁶¹

Dies sieht auch der Basler Ausschuss so: „PD estimates must be a long-run average of one-year default rates for borrowers in the grade“.⁶² Die weiteren Anforderungen⁶³ konkretisieren dann auch den Begriff „long-run“: Es müssen Beobachtungswerte über mindestens fünf Jahre vorhanden sein. Ist mehr relevantes Datenmaterial verfügbar,

⁵⁹Vgl. BCBS (2006, S. 99, Paragraph 446).

⁶⁰Vgl. z. B. Krink et al. (2008, S. 2275), Foglia et al. (2001, S. 424) oder Höse & Huschens (2003, S. 140).

⁶¹Vgl. Huschens & Locarek-Junge (2002) und Höse & Huschens (2003), welche ein Einfaktormodell für die Bonitätsvariablen mit einem Schwellenwertmodell für die Ausfallvariablen verbinden.

⁶²BCBS (2006, S. 99, Paragraph 447).

⁶³Vgl. BCBS (2006, S. 99, Paragraphen 461-467).

muss es verwendet werden. Unterstützt wird diese Anforderung durch eine Studie von Carey & Hrycay (2001, insb. S. 248-250). Damit ihre PD-Schätzwerte gewisse (sehr einfache) Qualitätskriterien⁶⁴ erfüllen, benötigen sie für ihren Datensatz Beobachtungen aus mehr als 10 Jahren. Carey & Hrycay sprechen aber auch an, dass nicht nur die Anzahl der Jahre eine Rolle spielt, sondern auch die Tatsache, dass sowohl wirtschaftlich gute als auch schlechte Jahre in die Schätzung einbezogen werden. Damit schließlich verlässliche Ausfallwahrscheinlichkeiten geschätzt werden, impliziert die Verwendung von Daten aus mehreren Perioden aber, dass das bankinterne Ratingsystem in den ersten beiden Dimensionen, d. h. bei der Bonitätsbeurteilung auf Kreditnehmerebene und bei der Ratingklasseneinteilung, im betreffenden Zeitraum unverändert geblieben ist.

Mit Hilfe der geschätzten Ausfallwahrscheinlichkeiten kann schließlich auch das bankinterne Ratingsystem validiert werden. Idealerweise besteht in diesem Zusammenhang der Wunsch nach beobachteten durchschnittlichen Ausfallquoten, die sich zwischen den Klassen statistisch unterscheiden. Mit einer hinreichend langen Zeitreihe an Ausfallraten, die mindestens einen kompletten Wirtschaftszyklus überdeckt, kann die Validierung gemäß Foglia et al. (2001, S. 442) wie folgt ausgeführt werden: Prüfe,

- ob jede jährliche Ausfallrate einer jeden Ratingklasse in einem Bereich liegt, der das Risiko dieser Ratingklasse eindeutig beschreibt, und
- ob die durchschnittlichen Ausfallquoten über die Ratingklassen monoton sind.

Etwas schärfer ausgedrückt werden nichtüberlappende und mit zunehmender Kreditqualität monoton fallende Ausfallquoten verlangt, d. h. formal

$$\max_t \bar{y}_{1t} \geq \min_t \bar{y}_{1t} > \max_t \bar{y}_{2t} \geq \min_t \bar{y}_{2t} > \dots > \max_t \bar{y}_{Kt} \geq \min_t \bar{y}_{Kt}. \quad (2.22)$$

Dabei bezeichnet \bar{y}_{kt} , $k = 1, \dots, K$, $t = 1, \dots, T$ die realisierte Ausfallquote in Ratingklasse k zum Zeitpunkt t . Ferner sei hier Ratingklasse 1 diejenige Klasse mit der niedrigsten Kreditqualität und umgekehrt Ratingklasse K diejenige mit der höchsten. Die Überschneidungsfreiheit der jährlichen Ausfallquoten ist jedoch bei Klassen mit sehr geringem Risiko nicht immer zu erreichen. Dies liegt in der Natur der Sache: In den besten Ratingklassen sind Ausfälle seltene Ereignisse, häufig gibt es in bestimmten Jahren überhaupt keine Ausfälle. So kann manchmal schon ein Ausfall die Ausfallquote der theoretisch besseren Ratingklasse über das Niveau der schlechteren Klasse heben.⁶⁵

⁶⁴„[A] goal of quantification is that estimates differ from the true value by no more than one-third no more than half the time (a rather loose standard of accuracy)“ (Carey & Hrycay, 2001, S. 248).

⁶⁵Vgl. Foglia et al. (2001, S. 453).

Kapitel 3

Bruchpunktschätzung

Eine zentrale Fragestellung aus dem vorangegangenen Kapitel ist die nach der Aufteilung von Kreditnehmern auf Ratingklassen: Wo sind die Trennstellen zwischen den Ratingklassen zu setzen? Anschließend stellt sich die Frage nach der Schätzung des Ausfallrisikos in allen Ratingklassen. Beschränkt man sich zunächst auf den Fall

- zweier Ratingklassen und
- stochastischer Unabhängigkeit der Kreditnehmer,

dann existieren geeignete Methoden im Bereich der Bruchpunktschätzung. Speziell beschäftigen wir uns mit den Arbeiten von

- Dempfle & Stute (2002),
- Ferger & Klotsche (2009) sowie
- Ferger (2009).

Die Auswertung der Literatur führt zur folgenden Gliederung dieses Kapitels. In Abschnitt 3.1 werden das Modell und seine Eigenschaften vorgestellt. Ferner wird das Schätzproblem einführend besprochen. In Abschnitt 3.2 wird mit dem Dempfle-Stute-Schätzer ein nichtparametrischer Schätzer für den gesuchten Bruchpunkt diskutiert. Der darauf folgende Abschnitt 3.3 beschäftigt sich mit dem Maximum-Likelihood-Schätzer für den Bruchpunkt. Dabei müssen jedoch gewisse Parameter des Modells als bekannt vorausgesetzt werden. Diese Schwachstelle wird in Abschnitt 3.4 behoben, indem diese beiden Ansätze zum sogenannten Zwei-Schritt-Plug-In-Schätzer kombiniert werden. Ein kurzer Vergleich der Schätzer in Abschnitt 3.5 rundet das Kapitel ab.

Die Ziele dieses Abschnitts sind dabei zum Teil widerstreitend. Zum einen soll versucht werden, die vorgeschlagenen Methoden und Ergebnisse verständlicher zu beschreiben, sie ausführlicher zu motivieren und technische Annahmen und Aussagen

nichttechnisch auszudrücken. Auf der anderen Seite geht es auch um eine detaillierte Analyse der Methoden und Ergebnisse, die Behebung von Unstimmigkeiten sowie die Diskussion von Zusammenhängen. Hierfür ist es teilweise unmöglich, auf technische Hilfsmittel zu verzichten. Dies führt dazu, dass insbesondere in den Abschnitten 3.2 bis 3.4 ein formeller Stil gepflegt wird.

3.1 Modell und Schätzproblem

Im folgenden Abschnitt werden das Modell und seine Eigenschaften beschrieben sowie das Schätzproblem vorgestellt. Da das Modell auch außerhalb des Kreditrisikos verwendbar ist, werden im restlichen Teil des Kapitels kontextabhängige Bezeichnungen weitestgehend vermieden.

3.1.1 Modellstruktur

Gegeben ist eine Gesamtheit von Merkmalsträgern. An jedem dieser Merkmalsträger werden zwei Merkmale X und Y beobachtet. Das binäre Merkmal Y gibt an, ob der Merkmalsträger eine interessierende Eigenschaft besitzt ($Y = 1$) oder ob er sie nicht besitzt ($Y = 0$). Im Folgenden werden dafür auch die Begriffe Erfolg bzw. Misserfolg verwendet.¹ Beobachtungen der Merkmale X und Y entstehen zufällig. Dementsprechend werden beide Merkmale als Zufallsvariable über einem gemeinsamen Wahrscheinlichkeitsraum modelliert. Die Zufallsvariablen X und Y sind durch folgende Modellstruktur verknüpft:

$$P(Y = y|X = x) = \begin{cases} \kappa(x), & \text{falls } y = 1, \\ 1 - \kappa(x), & \text{falls } y = 0. \end{cases} \quad (3.1)$$

Dabei sei

$$\kappa(x) = \begin{cases} a, & \text{falls } x \leq \theta, \\ b, & \text{falls } x > \theta, \end{cases} \quad (3.2)$$

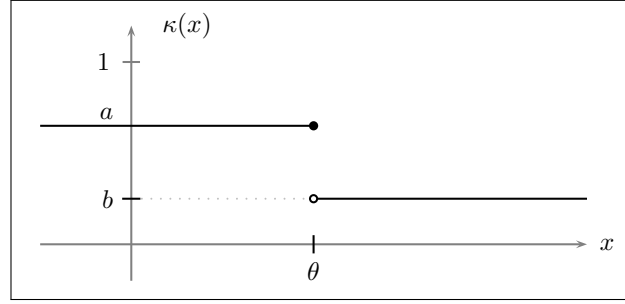
mit²

$$a \in]0, 1[, \quad b \in]0, 1[, \quad a \neq b \quad \text{und} \quad \theta \in \mathbb{R}.$$

In Abbildung 3.1 ist die Funktion κ für den Fall $a > b$ dargestellt. Die Parameter a und b werden als Risikoniveaus (*risk levels*) bezeichnet und sind unbekannt. Der

¹Die Begriffe Erfolg und Misserfolg enthalten keine inhaltliche Wertung. So kann Erfolg auch für eine negative Eigenschaft wie Krankheit oder Tod stehen.

²Die Bezeichnungen der Parameter wurden aus der Literatur übernommen. Üblicherweise werden Parameter mit griechischen Buchstaben bezeichnet.

Abbildung 3.1: Funktion κ für $a > b$

Parameter θ wird Bruchpunkt oder Schwellenwert (*split-point, change point, threshold*) genannt und ist ebenfalls unbekannt. Die unbekannte (Rand-)Verteilungsfunktion von X wird mit F bezeichnet. Sie ist unabhängig von den Parametern a , b und θ .

Für die Modellstruktur in (3.1) und für die Funktion κ in (3.2) existieren alternative Darstellungsmöglichkeiten, welche teilweise auch in den zugrunde liegenden Papieren von Dempfle & Stute (2002), Ferger & Klotsche (2009) sowie Ferger (2009) genutzt werden. Für den Zusammenhang in (3.1) gibt es beispielsweise folgende Schreibweisen:

- $P(Y = 1|X = x) = \kappa(x)$ und $P(Y = 0|X = x) = 1 - \kappa(x)$,
 - $Y|X = x \sim \text{Ber}(\kappa(x))$ oder
 - $P(Y \in \cdot | X = x) = B(1, \kappa(x))$.
- (3.3)

Dabei steht die Bezeichnung $\text{Ber}(p) = B(1, p)$ für eine Bernoulli-Verteilung mit Erfolgsparameter p . Die Bernoulli-Verteilung ist eine Zweipunktverteilung auf den Zahlen 0 und 1. Sie ist ein Spezialfall der Binomialverteilung, welche als $B(m, p)$ notiert wird. Für die Funktion κ in (3.2) ergeben sich mit Hilfe der Indikatorfunktion die folgenden geschlossenen Darstellungen:

$$\begin{aligned}
 \kappa(x) &= a\mathbb{1}_{]-\infty, \theta[}(x) + b\mathbb{1}_{] \theta, +\infty[}(x) \\
 &= a + (b - a)\mathbb{1}_{] \theta, +\infty[}(x) \\
 &= b + (a - b)\mathbb{1}_{]-\infty, \theta[}(x).
 \end{aligned}
 \tag{3.4}$$

Bei der gewählten Modellierung mit stochastischem Design³ kann man sich die Entstehung des Zufallsvektors (X, Y) als zweistufiges Zufallsexperiment vorstellen. Die Zufallsvariable X legt über die Funktion κ die Erfolgswahrscheinlichkeit der Bernoulli-verteilten Zufallsvariable Y fest. Anschließend kommt es zum Erfolg oder nicht. Die Zufallsvariable $\kappa(X)$ fungiert sozusagen als stochastischer Parameter der Bernoulli-

³Im Gegensatz zum stochastischen Design steht das Design als Experimentplanung. Dabei sind die Werte des Merkmals X nichtstochastisch, d. h. sie können vom Experimentator festgelegt werden.

Verteilung von Y . Die zufällige Erfolgswahrscheinlichkeit hat dabei die zwei Niveaus a und b .

In diesem Sinne ist das gegebene Modell ein Bernoulli-Mischungsmodell (*Bernoulli mixture model*). Die Wahrscheinlichkeitsverteilung von Y wird in diesem Zusammenhang auch als Mischverteilung (*mixed distribution*) oder zusammengesetzte Verteilung (*compound distribution*) bezeichnet. Da die Zufallsvariable $\kappa(X)$ nur die zwei verschiedenen Werte a und b annimmt, handelt es sich im vorliegenden Fall um eine endliche oder diskrete Mischung (*finite mixture*).⁴

Weil die Zufallsvariable Y nur die Werte 0 und 1 annehmen kann, gilt für den auf $X = x$ bedingten Erwartungswert von Y die Gleichung

$$\mathbb{E}[Y|X = x] = P(Y = 1|X = x) = \kappa(x), \quad x \in \mathbb{R}. \quad (3.5)$$

Damit ist $\kappa(x)$ die Regressionsfunktion 1. Art der Regression von Y auf X . Das vorliegende Modell lässt sich also auch als Regressionsmodell in der Form

$$Y = \kappa(X) + \epsilon \quad (3.6)$$

auffassen. Dabei ist $\kappa(X) = \mathbb{E}[Y|X]$ die bedingte Erwartung von Y gegeben X und $\epsilon := Y - \kappa(X)$ ist die Residualvariable oder Störgröße.⁵

Aufgrund von (3.5) und der Treppenform der Funktion κ ist das gegebene Modell durch eine abrupte Änderung des bedingten Erwartungswerts von Y gekennzeichnet, wenn die Eingangsvariable X den Schwellenwert θ überschreitet.⁶ Da sich beim Überschreiten des Bruchpunkts aber nur der bedingte Erwartungswert ändert, nicht aber das Ergebnis selbst, ist das hier gegebene Modell klar vom Schranken- oder Schwellenwertmodell (*threshold model*) zu unterscheiden. In diesem Modell tritt der Erfolg $Y = 1$ genau dann ein, wenn die zweite Zufallsvariable X einen kritischen Wert c (*threshold, cut-off level*) über- oder unterschreitet. Das Schwellenwertmodell ist ein typisches Modell der Kreditrisikomodellierung. Häufig ist dabei die Zufallsvariable X latent, d. h. unbeobachtbar.⁷

⁴Vgl. Müller (1991, S. 263), Bluhm et al. (2010, S. 55), Everitt (2006, S. 4880-4881) und Titterton (2006, S. 4890-4891).

⁵Vgl. Fenger & Klotsche (2009, S. 94) und Müller (1991, S. 326).

⁶Vgl. Dempfle & Stute (2002, S. 233).

⁷Vgl. Höse & Huschens (2011, Abschnitt 3.2), Frey et al. (2001, S. 111), McNeil et al. (2005, Abschnitt 8.3) oder Bluhm et al. (2010, S. 65).

Unser Modell (3.1) und (3.2) kann im soeben erwähnten Kontext des Kreditrisikos angewendet werden:

- Dabei sei X eine Bonitätsvariable, deren Verteilung unbekannt ist.
- Die Bernoulli-verteilte Zufallsvariable Y spiegelt das Ausfallverhalten des untersuchten Individuums wider. Dabei sei der Zustand 1 mit einem Ausfall gleichzusetzen. Der Zustand 0 steht für den Nichtausfall.
- Das Verhalten der Zufallsvariable Y ist über die Modellgleichungen (3.1) und (3.2) an die Ausprägung der Bonitätsvariable X geknüpft.
- Die Risikoniveaus a und b entsprechen in diesem Fall bedingten Ausfallwahrscheinlichkeiten.
- Der Bruchpunkt θ trennt die Bonitäten mit höherer und niedrigerer Ausfallwahrscheinlichkeit.

Neben dem Kreditrisiko ist eine Vielzahl weiterer Anwendungsbereiche zumindest denkbar. Eine Auswahl an Beispielen aus Medizin, Natur- und Wirtschaftswissenschaften liefert Tabelle 3.1.

Merkmal X	binäres Merkmal Y
Blutdruck	Tod (Sterblichkeit) ^a
Blutzucker	Krebs ^b
Fettleibigkeit (Taillenumfang, Body-Mass-Index)	Herz-Kreislauf-Erkrankungen ^c
Pupillendurchmesser	Progressive supranukleäre Blickparese ^d
radioaktive Strahlung	Krebs, Missbildungen bei Nachkommen
Konzentration von Umweltgiften	<ul style="list-style-type: none"> • Krankheiten^e • Artenvorkommen^f
Jahresmitteltemperatur	Artenvorkommen
Arbeitslosenquote	Zuwanderung/Abwanderung ^g
Bonität	Ausfallverhalten

^a Vgl. Berry et al. (2012).

^b Vgl. Stocks et al. (2009).

^c Vgl. Ferger & Klotsche (2009).

^d Vgl. Schmidt et al. (2007).

^e Vgl. Kuhlisch (2003).

^f Vgl. Banerjee & McKeague (2007).

^g Vgl. Schlömer & Bucher (2001) sowie Abschnitt 5.2 dieser Arbeit.

Tabelle 3.1: Denkbare und untersuchte Anwendungsbeispiele des Modells

3.1.2 Eigenschaften des Modells

Im Folgenden wird das gegebene Modell näher untersucht. Weitere Modelleigenschaften werden angegeben. Aus (3.1) und (3.2) folgt z. B.

$$P(Y = 1|X = x) = \begin{cases} a & \text{für alle } x \leq \theta, \\ b & \text{für alle } x > \theta. \end{cases}$$

Daher gilt⁸

$$a = P(Y = 1|X \leq \theta) \quad \text{und} \quad b = P(Y = 1|X > \theta).$$

Wie bereits erwähnt, gilt für den auf $X = x$ bedingten Erwartungswert von Y die Gleichung

$$\mathbb{E}[Y|X = x] = \kappa(x), \quad x \in \mathbb{R}.$$

Für die auf $X = x$ bedingte Varianz von Y folgt aus (3.3)

$$\mathbb{V}[Y|X = x] = \kappa(x)(1 - \kappa(x)), \quad x \in \mathbb{R}. \quad (3.7)$$

Die Zufallsvariable $\kappa(X)$ besitzt eine Zweipunktverteilung auf den Zahlen a und b . Es gilt

$$\kappa(X) = a \Leftrightarrow X \leq \theta \quad \text{und} \quad \kappa(X) = b \Leftrightarrow X > \theta.$$

Somit ist

$$P(\kappa(X) = a) = P(X \leq \theta) = F(\theta) \quad \text{und} \quad P(\kappa(X) = b) = P(X > \theta) = 1 - F(\theta).$$

Für den Erwartungswert der Zufallsvariable $\kappa(X)$ folgt daher

$$\mathbb{E}[\kappa(X)] = aF(\theta) + b(1 - F(\theta)).$$

Diese Konvexkombination von a und b ist gleichzeitig die unbedingte Erfolgswahrscheinlichkeit der Bernoulli-Variable Y , d. h.

$$E[Y] = P(Y = 1) = aF(\theta) + b(1 - F(\theta)) \in]0, 1[. \quad (3.8)$$

⁸Vgl. Ferger & Klotsche (2009, S. 94).

Das Ergebnis erhält man auf verschiedenen Wegen. Eine Möglichkeit wäre die Anwendung des Gesetzes vom iterierten Erwartungswert⁹

$$\mathbb{E}[Y] = \mathbb{E}[\mathbb{E}[Y|X]] = \mathbb{E}[\kappa(X)],$$

eine weitere die Verwendung der Eigenschaften eines endlichen Bernoulli-Mischungsmodells.¹⁰ Weil Y Bernoulli-verteilt ist und daher wegen $Y \in \{0, 1\}$ die Gleichung $Y^2 = Y$ gilt, folgt für die Varianz

$$\begin{aligned} \mathbb{V}[Y] &= \mathbb{E}[Y^2] - (\mathbb{E}[Y])^2 = \mathbb{E}[Y] - (\mathbb{E}[Y])^2 = \mathbb{E}[Y] (1 - \mathbb{E}[Y]) \\ &= (aF(\theta) + b(1 - F(\theta))) (1 - aF(\theta) - b(1 - F(\theta))) \\ &= (b + (a - b)F(\theta)) (1 - b - (a - b)F(\theta)) \\ &= b(1 - b) + (1 - 2b)(a - b)F(\theta) - (a - b)^2 F(\theta)^2. \end{aligned}$$

Einige Bemerkungen zu den Eigenschaften der Störgröße $\epsilon = Y - \kappa(X)$ aus dem Regressionsmodell (3.6) sollen diesen Abschnitt abschließen. Die Störgröße ist bedingt zentriert, d. h.¹¹

$$\mathbb{E}[\epsilon|X] = \mathbb{E}[Y - \kappa(X)|X] = \mathbb{E}[Y|X] - \kappa(X) = \kappa(X) - \kappa(X) = 0 \text{ f. s.}, \quad (3.9)$$

und folglich zentriert:¹²

$$\mathbb{E}[\epsilon] = \mathbb{E}[\mathbb{E}[\epsilon|X]] = 0.$$

Unter der zusätzlichen Voraussetzung, dass $\mathbb{E}[\epsilon X]$ und $\mathbb{E}[X]$ existieren und endlich sind, gilt für die Kovarianz von ϵ und X die Gleichung¹³

$$\text{Cov}[\epsilon, X] = \mathbb{E}[\epsilon X] - \underbrace{\mathbb{E}[\epsilon]}_{=0} \mathbb{E}[X] = \mathbb{E}[\mathbb{E}[\epsilon X|X]] = \mathbb{E}[X \underbrace{\mathbb{E}[\epsilon|X]}_{=0 \text{ f. s.}}] = 0,$$

d. h. die Störvariable ϵ und der Regressor X sind unkorreliert. Sie sind aber nicht stochastisch unabhängig, denn die bedingte Varianz von ϵ ist von X abhängig. Es gilt¹⁴

$$\mathbb{V}[\epsilon|X] = \kappa(X)(1 - \kappa(X)) \text{ f. s.},$$

⁹Vgl. z. B. Karr (1993, S. 226, Proposition 8.13).

¹⁰Vgl. Müller (1991, S. 263, Z. –5). Siehe auch Bluhm et al. (2010, S. 55, Gleichung (2.3)).

¹¹Vgl. zum Ergebnis auch Ferger & Klotsche (2009, S. 94) und Dempfle & Stute (2002, S. 234), zum Lösungsweg z. B. Karr (1993, S. 232-233, Proposition 8.27 und Theorem 8.34).

¹²Vgl. zum Gesetz vom iterierten Erwartungswert z. B. Karr (1993, S. 226, Proposition 8.13).

¹³Zum vorletzten Gleichheitszeichen vgl. z. B. Gänssler & Stute (1977, S. 189, Satz 5.2.7).

¹⁴Vgl. auch Ferger & Klotsche (2009, S. 94) und Dempfle & Stute (2002, S. 234).

denn wegen (3.9) und $Y \in \{0, 1\}$ ist¹⁵

$$\begin{aligned}
 \mathbb{V}[\epsilon|X] &= \mathbb{E}[\epsilon^2|X] - \underbrace{(\mathbb{E}[\epsilon|X])^2}_{=0 \text{ f. s.}} \\
 &= \mathbb{E}[(Y - \kappa(X))^2|X] \\
 &= \mathbb{E}[\underbrace{Y^2}_{=Y} - 2Y\kappa(X) + \kappa(X)^2|X] \\
 &= \underbrace{\mathbb{E}[Y|X]}_{=\kappa(X)} - 2\kappa(X) \underbrace{\mathbb{E}[Y|X]}_{=\kappa(X)} + \kappa(X)^2 \\
 &= \kappa(X) - \kappa(X)^2 \\
 &= \kappa(X)(1 - \kappa(X)) \text{ f. s.}
 \end{aligned}$$

3.1.3 Schätzproblem

Es gilt, die unbekannten Parameter unseres Modells (3.1) und (3.2), d. h. a , b und θ , zu schätzen, wobei die Priorität zunächst auf dem Bruchpunkt θ liegen soll. Dazu betrachten wir i.i.d.-Kopien des Zufallsvektors (X, Y) , d. h.

$$(X_i, Y_i) \stackrel{\text{i.i.d.}}{\sim} (X, Y), \quad i = 1, \dots, n.$$

Wie bereits erwähnt, unterstellen wir stochastisches Design und die Beobachtbarkeit der beiden Zufallsvariablen X und Y . Weil die Streuungsfunktion $\mathbb{V}[\epsilon|X = x] = \kappa(x)(1 - \kappa(x))$ nicht konstant ist¹⁶ und daher ϵ und X nicht stochastisch unabhängig sind, empfehlen Dempfle & Stute (2002, S. 234), bei der Schätzung des Bruchpunkts von der Verwendung der Residualvariablen ϵ_i Abstand zu nehmen. Sie schlagen stattdessen einen nichtparametrischen Schätzer für θ vor, der unter schwachen Voraussetzungen wünschenswerte Eigenschaften hat. Details dazu folgen im nächsten Abschnitt.

An dieser Stelle soll es vielmehr um eine Mindestvoraussetzung für die sinnvolle Schätzung von θ gehen. Der unbekannte Parameter θ ist der Trennpunkt zwischen höherer und niedrigerer bedingter Erfolgswahrscheinlichkeit. Um diese Modellstruktur bei der Schätzung auszunutzen, benötigt man sowohl links als auch rechts von θ Beobachtungen des Merkmals X . Eine Mindestvoraussetzung dafür ist, dass die Zufallsvariable X beiderseits von θ Wahrscheinlichkeitsmasse besitzt. Mit anderen Worten heißt das, dass $0 < F(\theta) < 1$ gelten muss. Dies setzen wir im Folgenden stets voraus. Wäre die Voraussetzung nicht erfüllt, d. h. würde $P(X \leq \theta) = F(\theta) = 1$ oder

¹⁵Vgl. zum Lösungsweg auch Karr (1993, S. 232-233, Proposition 8.27 und Theorem 8.34).

¹⁶Man spricht in diesem Fall von Heteroskedastizität.

$P(X \leq \theta) = F(\theta) = 0$ gelten, dann lägen fast sicher alle X -Beobachtungen links bzw. rechts von der Trennstelle θ . Folglich wäre der Parameter θ nicht schätzbar.¹⁷

Erste Arbeiten auf dem Gebiet der Bruchpunktschätzung stammen von Hinkley (1969, 1971). Die weitere Entwicklung der nichtparametrischen Schätzung skizzieren Dempfle & Stute (2002, S. 233-234). Sie gehen dabei insbesondere auf die Annahmen und Hauptarbeitsschritte ein, die die vorgeschlagenen „Smoothing“-Ansätze in vielen praktischen Anwendungen als weniger gut geeignet erscheinen lassen. Zum Beispiel wird typischerweise vorausgesetzt, dass die Residualvariablen ϵ_i unabhängig von den erklärenden Variablen X_i und homoskedastisch sind, d. h. ihre Varianz über alle Beobachtungen konstant ist. Ferner sei die Wahl der Glättungsparameter nicht unproblematisch.¹⁸ Einen ähnlichen Abriss der Geschichte des Bruchpunktschätzproblems findet man auch in Ferger & Klotsche (2009, S. 96-97). Sie gehen dabei intensiver auf den Fall stochastischen Designs ein.

Ist schließlich ein geeigneter Schätzer für den Bruchpunkt θ gefunden, kann man davon ausgehend Schätzer für die Risikoniveaus a und b finden. Dies erscheint einfach, denn es bieten sich die Erfolgsquoten in den zwei durch den Bruchpunktschätzwert getrennten Gruppen an. Anschließend kann der Bruchpunktschätzer auf Basis der Risikoniveauschätzer revidiert werden. In den folgenden Abschnitten zeigt sich, dass die oben angeführte Literatur genau diesen Weg gegangen ist.

3.2 Dempfle-Stute-Schätzer

In diesem Abschnitt wird der von Dempfle & Stute (2002) vorgeschlagene Bruchpunktschätzer näher betrachtet. Dazu werden zunächst zwei Ideen präsentiert, die den Schätzer motivieren. Anschließend werden äquivalente Schreibweisen für den Schätzer dargestellt und Eigenschaften des zugrunde liegenden empirischen Prozesses beleuchtet. Danach wird auf verwandte Schätzer eingegangen. Abschließend werden Resultate zu den asymptotischen Eigenschaften der Bruchpunktschätzer vorgestellt.

3.2.1 Einführung

Im Folgenden werden zwei Ideen beschrieben, die zur Schätzung von θ führen.

1. Der erste Ansatz ist durch die in Abschnitt 2.1.2 getätigten Überlegungen zum Kreditrisiko motiviert. Als Schätzer für den unbekannten Parameter θ soll die Maximalstelle zum ROC-gap fungieren.

¹⁷Die Bedingung $0 < F(\theta) < 1$ und ihre Erläuterung findet man auch in Dempfle & Stute (2002, S. 235). Um Güteeigenschaften wie Konsistenz oder n -Konsistenz nachzuweisen, muss die Mindestvoraussetzung weiter verschärft werden. Details folgen in den nächsten Abschnitten dieser Arbeit.

¹⁸Vgl. zu diesem Problem auch die Ausführungen in Abschnitt 4.1 der vorliegenden Arbeit.

2. Für den zweiten Ansatz werden zunächst Teilmengen der Stichprobe gebildet. Die Einteilung der Teilstichproben erfolgt dabei anhand der Größe des Merkmals X . Anschließend wird für jede Teilstichprobe die tatsächliche Anzahl der Erfolge mit dem natürlichen Schätzer für diese Anzahl verglichen.

Es wird sich zeigen, dass beide Ansätze zum selben Ergebnis führen.

Im ersten Ansatz dient die Receiver Operating Characteristic (ROC) als Grundlage. Die ROC dient dem Vergleich verschiedener Scoringssysteme. Weil nicht alle Scoringmodelle mit Hilfe der ROC vergleichbar sind, werden eindimensionale Maße wie das ROC-gap definiert. Das ROC-gap misst den maximalen vertikalen Abstand zwischen der ROC des betrachteten Systems und der trivialen ROC¹⁹, d. h.

$$\text{ROC-gap} = \max_{x \in \mathbb{R}} d^{\text{ROC}}(x),$$

wobei

$$d^{\text{ROC}}(x) = HR(x) - FAR(x), \quad x \in \mathbb{R}.$$

Um die Trennstelle θ zwischen der höheren und der niedrigeren bedingten Erfolgswahrscheinlichkeit (im Kontext des Kreditrisikos: Ausfallwahrscheinlichkeit) zu schätzen, lautet die Idee: Verwende als Schätzwert für θ eine Maximalstelle der Funktion d^{ROC} , d. h.

$$\hat{\theta}_n := \operatorname{argmax}_{x \in \mathbb{R}} d^{\text{ROC}}(x). \quad (3.10)$$

Eine genauere Beschreibung des argmax -Funktional wird in Abschnitt 3.2.2, Bemerkung 3.9, nachgeliefert.

Zur Hinführung zur zweiten Idee betrachten wir zunächst folgendes Problem: Gegeben sei eine Gesamtheit von n Individuen. Das binäre Merkmal Y gebe an, ob das Individuum eine interessierende Eigenschaft besitzt ($Y = 1$, Erfolg) oder ob es sie nicht besitzt ($Y = 0$, Misserfolg). Des Weiteren betrachten wir eine beliebige Teilmenge aus der Gesamtpopulation. Wir kennen sowohl die Anzahl der Erfolge in der Gesamtpopulation, z. B. 20, als auch den Anteil der Teilpopulation an der Gesamtpopulation, beispielsweise ein Viertel. Die Frage lautet: Wie viele Erfolge erwartet man in der Teilpopulation?

Die naive Antwort ist: „Das Produkt der beiden Größen, also 5!“ Sie beruht auf der Annahme, dass die Teilpopulation keinen Einfluss auf die Erfolgswahrscheinlichkeit hat. Das ist häufig nicht der Fall. Insbesondere ist das dann nicht der Fall, wenn die Teilpopulation anhand eines weiteren Merkmals X gebildet wird, welches Einfluss

¹⁹Eine ausführlichere Diskussion der genannten Konzepte ist in Abschnitt 2.1.2 zu finden.

auf das Merkmal Y , oder genauer gesagt auf dessen Erfolgswahrscheinlichkeit, hat. Wird z. B. in der Patientenschaft eines Allgemeinmediziners die Teilmenge aller Patienten mit Bluthochdruck betrachtet, so hat diese ein höheres Herzinfarkt- und damit Sterblichkeitsrisiko als die Menge der Patienten mit normalem Blutdruck. Für weitere Beispiele sei auf Tabelle 3.1 verwiesen.

Um den Bruchpunkt θ aus einer Stichprobe vom Umfang n zu schätzen, soll nun die tatsächliche Anzahl der Erfolge in gewissen Teilstichproben mit der naiven Überlegung verglichen werden. Dabei werden Teilstichproben der Form

$$\{i = 1, \dots, n | X_i \leq x\}$$

betrachtet. Die Schranke x wird hierbei vorgegeben. Formal ergeben sich dann folgende Ausdrücke für die relevanten Größen:

- für die Anzahl der Erfolge in der Gesamtstichprobe

$$\sum_{i=1}^n Y_i,$$

- für die Anzahl der Individuen in der Teilstichprobe

$$\sum_{i=1}^n \mathbb{1}_{\{X_i \leq x\}},$$

- daher für den Anteil der Teilstichprobe an der Gesamtstichprobe

$$\frac{\sum_{i=1}^n \mathbb{1}_{\{X_i \leq x\}}}{n} = \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{\{X_i \leq x\}} = F_n(x),$$

- somit für den naiven Schätzer für die Anzahl der Erfolge in der Teilstichprobe

$$F_n(x) \sum_{i=1}^n Y_i$$

- und schließlich für die tatsächliche Anzahl der Erfolge in der Teilstichprobe

$$\sum_{i=1}^n \mathbb{1}_{\{X_i \leq x\}} Y_i.$$

Nehmen wir nun an, dass die bedingte Erfolgswahrscheinlichkeit bei kleinen Werten von X größer ist als die für große Werte von X . Das heißt in unserem Modell, dass $a > b$ gilt. Dann erwarten wir aufgrund der zunächst größeren bedingten Erfolgs-

wahrscheinlichkeit a , dass die tatsächliche Anzahl der Erfolge in der Teilstichprobe mit zunehmendem x zunächst schneller steigt als die naiv erwartete Anzahl. Die Differenz der Größen, d. h.

$$D_n(x) := \sum_{i=1}^n \mathbb{1}_{\{X_i \leq x\}} Y_i - F_n(x) \sum_{i=1}^n Y_i, \quad x \in \mathbb{R}, \quad (3.11)$$

sollte mit zunehmenden x also zunächst wachsen. Überschreitet x den Bruchpunkt θ , kommen aufgrund der nun geringeren bedingten Erfolgswahrscheinlichkeit b weniger Erfolge hinzu als naiv erwartet. Die tatsächliche Anzahl der Erfolge in der Teilstichprobe und die naive Erwartung gleichen sich wieder an. Oder anders ausgedrückt: Die Differenz in (3.11) wird wieder kleiner.

Eine einleuchtende Möglichkeit, die Trennstelle θ zu schätzen, ist es also, diejenige Stelle als Bruchpunkt anzunehmen, bei der die Differenz D_n am größten wird. Weil die Schreibweise in (3.11) zwar gut zu interpretieren, jedoch formal unhandlich ist, gilt es, hier noch eine Vereinfachung zu schaffen. Dazu führen wir die zufällige markierte empirische Verteilungsfunktion²⁰

$$H_n(x) := \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{\{X_i \leq x\}} Y_i, \quad x \in \mathbb{R} \quad (3.12)$$

ein, multiplizieren die Differenzenfunktion D_n mit dem Faktor $\frac{1}{n}$ und erhalten

$$\begin{aligned} \frac{1}{n} D_n(x) &= \frac{1}{n} \left(\sum_{i=1}^n \mathbb{1}_{\{X_i \leq x\}} Y_i - F_n(x) \sum_{i=1}^n Y_i \right) \\ &= \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{\{X_i \leq x\}} Y_i - F_n(x) \frac{1}{n} \sum_{i=1}^n Y_i \\ &= H_n(x) - \bar{Y} F_n(x), \quad x \in \mathbb{R}. \end{aligned}$$

Durch die Multiplikation mit der positiven Konstante $\frac{1}{n}$ ändert sich zwar das Maximum, nicht aber die Maximalstelle. Daher führt der zweite Ansatz zu dem in der folgenden Definition beschriebenen Schätzer für den Parameter θ .

Definition 3.1 (Dempfle-Stute-Schätzer)

Im durch die Gleichungen (3.1) und (3.2) beschriebenen Modell heißt der Schätzer

$$\hat{\theta}_n := \operatorname{argmax}_{x \in \mathbb{R}} S_n(x), \quad (3.13)$$

²⁰Diese unterscheidet sich von der empirischen Verteilungsfunktion F_n durch den Faktor Y_i in jedem Summanden. Wegen $Y_i \in \{0, 1\}$ werden so die Erfolge ($Y_i = 1$) markiert.

wobei

$$S_n(x) := H_n(x) - \bar{Y}F_n(x), \quad x \in \mathbb{R}, \quad (3.14)$$

Dempfle-Stute-Schätzer²¹ für den Parameter θ .

Bemerkung 3.2 (Zur Definition von $\hat{\theta}_n$)

- (i) In beiden vorgestellten Ansätzen (3.10) und (3.13) wird der Schätzer für θ mit $\hat{\theta}_n$ bezeichnet. Das ist formal korrekt, weil die Schätzer – wie sich im Folgenden zeigen wird – gleich sind.
- (ii) Eine genauere Beschreibung des *argmax*-Funktionals wird in Abschnitt 3.2.2, Bemerkung 3.9, nachgeliefert. Bis dahin soll die Vorstellung genügen, dass es sich bei $\hat{\theta}_n$ um eine Maximalstelle der betrachteten Funktion S_n handelt.

In der Literatur wird zur Motivation des Schätzers $\hat{\theta}_n$ der Grenzprozess zu S_n betrachtet. Dieser Grenzprozess S besitzt unter den Modellvoraussetzungen eine Dreiecksform. Im Fall $a > b$ wächst er bis zur Stelle θ . Anschließend fällt er. Somit ist der unbekannte Parameter θ die Maximalstelle von S . Um diese Maximalstelle zu schätzen, wird die Maximalstelle $\hat{\theta}_n$ des empirischen Prozesses S_n verwendet.

3.2.2 Äquivalente Schreibweisen

In diesem Abschnitt wird zunächst gezeigt, dass die beiden vorgestellten Ansätze (3.10) und (3.13) denselben Schätzer für θ liefern. Anschließend wird auf die äquivalenten Definitionen des Schätzers in der Literatur eingegangen.

Den Hinweis auf Ersteres, d. h. die Gleichheit der Ansätze (3.10) und (3.13), liefert ein Blick auf Abbildung 3.2. Dort ist die in Definition 3.1 eingeführte Funktion S_n für die bereits in Abschnitt 2.1.2 verwendeten Beispieldatensätze A und B dargestellt. Beim Vergleich mit Abbildung 2.3 (siehe S. 18) fällt die gleiche Form der empirischen Prozesse S_n und d^{ROC} sofort ins Auge. Der funktionale Zusammenhang zwischen den beiden Prozessen wird im Folgenden näher untersucht.

Dazu klären wir zunächst die Zusammenhänge zwischen der zufälligen markierten empirischen Verteilungsfunktion H_n und den in Abschnitt 2.1.2 eingeführten empirischen Verteilungsfunktionen der Scores der Ausfälle HR bzw. Nichtausfälle FAR .

²¹Die Namenswahl folgt der Sprechweise von Ferger (2009) sowie Ferger & Klotsche (2009). Begründet wird sie durch die Arbeit von Dempfle & Stute (2002), die diesen Schätzer untersucht. Eine weitere Vernetzung der genannten Arbeiten mit der hier vorliegenden erfolgt im späteren Verlauf dieser Arbeit.

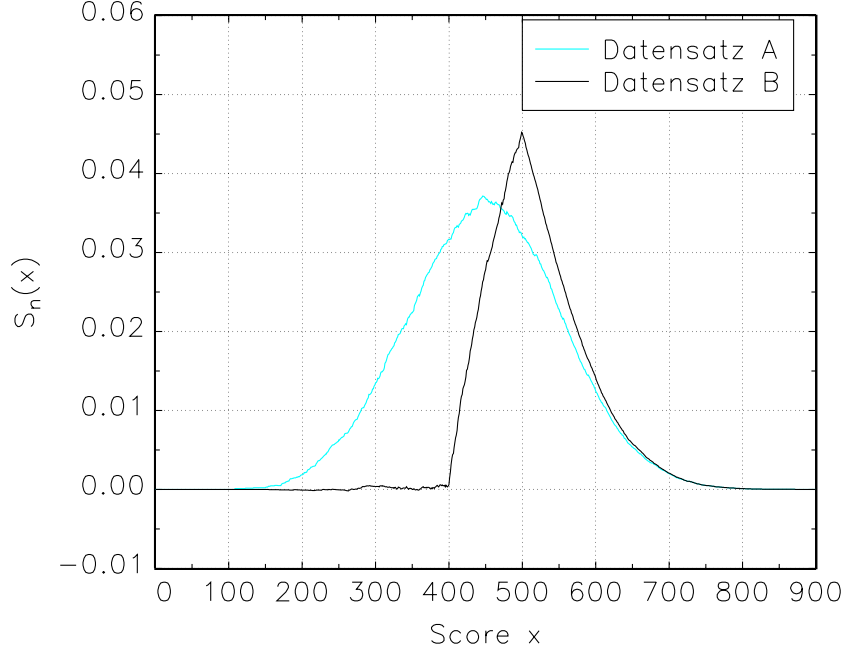


Abbildung 3.2: Empirischer Prozess S_n für Datensätze A und B

Lemma 3.3 Sei $Y_i \in \{0, 1\}$, $i = 1, \dots, n$ und $0 < \bar{Y} < 1$. Für alle $x \in \mathbb{R}$ gelten dann

$$H_n(x) = \bar{Y} HR(x) \quad (3.15)$$

und

$$F_n(x) - H_n(x) = (1 - \bar{Y}) FAR(x). \quad (3.16)$$

Beweis: Wegen $0 < \bar{Y} < 1$ gilt sowohl $N_D > 0$ als auch $N_{ND} > 0$. Daher sind die Funktionen HR und FAR definiert. Aufgrund der Definitionen in (3.12), (2.2) und (2.3) gilt

$$H_n(x) = \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{\{X_i \leq x\}} Y_i = \frac{N_D}{n} \frac{1}{N_D} \sum_{i=1}^n \mathbb{1}_{\{X_i \leq x\}} Y_i = \bar{Y} HR(x).$$

für alle $x \in \mathbb{R}$. Damit ist (3.15) bewiesen.

Die Gleichung (3.16) folgt nun aus Lemma 2.1 und (3.15). Für alle $x \in \mathbb{R}$ gilt

$$F_n(x) - H_n(x) = (\bar{Y} HR(x) + (1 - \bar{Y}) FAR(x)) - \bar{Y} HR(x) = (1 - \bar{Y}) FAR(x).$$

Damit ist der Beweis vollständig. \square

Im folgenden Satz wird nun der Zusammenhang zwischen den empirischen Prozessen S_n und d^{ROC} geklärt. Es zeigt sich, dass S_n und die Abstandsfunktion d^{ROC}

proportional sind. Der Proportionalitätsfaktor ist das Produkt von Erfolgsquote \bar{Y} und Misserfolgsquote $1 - \bar{Y}$.

Satz 3.4 *Sei $Y_i \in \{0, 1\}$, $i = 1, \dots, n$ und $0 < \bar{Y} < 1$. Für alle $x \in \mathbb{R}$ gilt dann*

$$S_n(x) = \bar{Y}(1 - \bar{Y})d^{\text{ROC}}(x). \quad (3.17)$$

Beweis: Wegen $0 < \bar{Y} < 1$ gilt sowohl $N_D > 0$ als auch $N_{ND} > 0$. Daher sind die Funktionen HR und FAR und damit gemäß (2.9) auch d^{ROC} definiert. Mit Lemma 3.3 sowie den Definitionen in (2.9) und (3.14) folgt

$$\begin{aligned} \bar{Y}(1 - \bar{Y})d^{\text{ROC}}(x) &= \bar{Y}(1 - \bar{Y})(HR(x) - FAR(x)) \\ &= (1 - \bar{Y})\bar{Y}HR(x) - \bar{Y}(1 - \bar{Y})FAR(x) \\ &= (1 - \bar{Y})H_n(x) - \bar{Y}(F_n(x) - H_n(x)) \\ &= H_n(x) - \bar{Y}H_n(x) - \bar{Y}F_n(x) + \bar{Y}H_n(x) \\ &= H_n(x) - \bar{Y}F_n(x) \\ &= S_n(x) \end{aligned}$$

für alle $x \in \mathbb{R}$. Damit ist die Behauptung bewiesen. \square

Folglich zeigt sich, dass die beiden vorgestellten Ansätze zur Schätzung des Bruchpunkts äquivalent sind, falls die Erfolgsquote \bar{Y} nicht null und ungleich eins ist.

Folgerung 3.5 *Sei $Y_i \in \{0, 1\}$, $i = 1, \dots, n$ und $0 < \bar{Y} < 1$. Dann gilt*

$$\hat{\theta}_n = \operatorname{argmax}_{x \in \mathbb{R}} S_n(x) = \operatorname{argmax}_{x \in \mathbb{R}} d^{\text{ROC}}(x).$$

Beweis: Wegen $0 < \bar{Y} < 1$ gilt sowohl $N_D > 0$ als auch $N_{ND} > 0$. Daher sind die Funktionen HR und FAR und damit gemäß (2.9) auch d^{ROC} definiert. Mit Satz 3.4 folgt

$$\hat{\theta}_n = \operatorname{argmax}_{x \in \mathbb{R}} S_n(x) = \operatorname{argmax}_{x \in \mathbb{R}} \bar{Y}(1 - \bar{Y})d^{\text{ROC}}(x) = \operatorname{argmax}_{x \in \mathbb{R}} d^{\text{ROC}}(x),$$

weil für die Bestimmung der Maximalstelle positive, nicht von x abhängige Faktoren wie \bar{Y} und $1 - \bar{Y}$ keine Rolle spielen. Damit ist die Behauptung bewiesen. \square

Bemerkung 3.6 (Zu Voraussetzung und Konsequenzen von Folgerung 3.5)

- (i) *Die Bedingung $0 < \bar{Y} < 1$ ist für schließlich alle $n \in \mathbb{N}$ fast sicher erfüllt, da die Erfolgsquote \bar{Y} gemäß dem starken Gesetz der großen Zahlen fast sicher gegen den Erwartungswert $\mathbb{E}[Y]$ konvergiert, welcher sich laut (3.8) im offenen Intervall $]0, 1[$ befindet.*

(ii) Die Folgerung zeigt, dass der Dempfle-Stute-Schätzer in unserem Modellrahmen gleich der Maximalstelle zum ROC-gap ist. Falls zusätzlich die Bedingung (2.11) gilt, ist somit der Dempfle-Stute-Schätzer auch die Maximalstelle der Teststatistik des Kolmogoroff-Smirnoff-Tests für das Hypothesenpaar in (2.10).²²

Kommen wir nun zu den weiteren Schreibweisen aus der Literatur. In der Arbeit von Ferger & Klotsche (2009, S. 94, Gleichung (1.4)) wird der Dempfle-Stute-Schätzer wie folgt eingeführt:

$$\hat{\theta}_n = \operatorname{argmax}_{x \in \mathbb{R}} \sum_{i=1}^n \sum_{j=1}^n (Y_i - Y_j) \mathbb{1}_{\{X_i \leq x < X_j\}}. \quad (3.18)$$

In Ferger (2009, S. 9, 53, 59) findet man die Darstellung

$$\hat{\theta}_n = \operatorname{argmax}_{x \in \mathbb{R}} \frac{1}{n(n-1)} \sum_{i=1}^n \sum_{\substack{j=1 \\ i \neq j}}^n (Y_i - Y_j) \mathbb{1}_{\{X_i \leq x < X_j\}}. \quad (3.19)$$

Diese Schreibweisen unterscheiden sich untereinander zum einen durch den Faktor $\frac{1}{n(n-1)}$ und zum anderen durch den Ausschluss der Summanden mit $i = j$ in (3.19). Da diese Summanden jedoch gleich null sind, sind die Doppelsummen trotzdem gleich. Weil – wie bereits im Beweis von Folgerung 3.5 festgestellt – positive, nicht von x abhängige Faktoren bei der Maximalstellensuche keine Rolle spielen, sind die Schätzer in (3.18) und (3.19) gleich. Dass sie ebenfalls äquivalent zu Definition 3.1 sind, gilt wegen

$$S_n(x) = n^{-2} \sum_{i=1}^n \sum_{j=1}^n (Y_i - Y_j) \mathbb{1}_{\{X_i \leq x < X_j\}}, \quad (3.20)$$

wie die folgende Bemerkung zeigt.²³ Der Faktor n^{-2} ist positiv und nicht von x abhängig. Er ist bei der Maximalstellenbestimmung also irrelevant.

Bemerkung 3.7 (Zum Beweis von (3.20))

Aus Symmetriegründen gilt

$$\begin{aligned} R := R(x) &:= \sum_{i=1}^n \sum_{j=1}^n (Y_i - Y_j) \mathbb{1}_{\{X_i \leq x\}} \mathbb{1}_{\{X_j \leq x\}} \\ &= \sum_{i=1}^n \sum_{j=1}^n Y_i \mathbb{1}_{\{X_i \leq x\}} \mathbb{1}_{\{X_j \leq x\}} - \sum_{i=1}^n \sum_{j=1}^n Y_j \mathbb{1}_{\{X_i \leq x\}} \mathbb{1}_{\{X_j \leq x\}} \\ &= 0 \end{aligned}$$

²²Siehe dazu S. 17.

²³Die „einfache Umformung“ wird in Ferger (2009, S. 59) in zwei Zeilen angedeutet. Hier wird sie ausführlich gezeigt.

für alle $x \in \mathbb{R}$. Damit und wegen²⁴ $\mathbb{1}_A = 1 - \mathbb{1}_{\bar{A}}$ sowie $\mathbb{1}_A \mathbb{1}_B = \mathbb{1}_{A \cap B}$ folgt für die Doppelsumme aus (3.18) und (3.19) die Gleichung

$$\begin{aligned}
& \sum_{i=1}^n \sum_{j=1}^n (Y_i - Y_j) \mathbb{1}_{\{X_i \leq x < X_j\}} \\
&= \sum_{i=1}^n \sum_{j=1}^n (Y_i - Y_j) \mathbb{1}_{\{X_i \leq x\}} \mathbb{1}_{\{X_j > x\}} \\
&= \sum_{i=1}^n \sum_{j=1}^n (Y_i - Y_j) \mathbb{1}_{\{X_i \leq x\}} \mathbb{1}_{\{X_j > x\}} + \underbrace{\sum_{i=1}^n \sum_{j=1}^n (Y_i - Y_j) \mathbb{1}_{\{X_i \leq x\}} \mathbb{1}_{\{X_j \leq x\}}}_{=R=0} \\
&= \sum_{i=1}^n \sum_{j=1}^n (Y_i - Y_j) \mathbb{1}_{\{X_i \leq x\}} \underbrace{(\mathbb{1}_{\{X_j > x\}} + \mathbb{1}_{\{X_j \leq x\}})}_{=1} \\
&= \sum_{i=1}^n \sum_{j=1}^n (Y_i - Y_j) \mathbb{1}_{\{X_i \leq x\}} \\
&= \sum_{i=1}^n \mathbb{1}_{\{X_i \leq x\}} \sum_{j=1}^n (Y_i - Y_j) \\
&= \sum_{i=1}^n \mathbb{1}_{\{X_i \leq x\}} \left(nY_i - \sum_{j=1}^n Y_j \right) \\
&= n \sum_{i=1}^n \mathbb{1}_{\{X_i \leq x\}} \left(Y_i - \frac{1}{n} \sum_{j=1}^n Y_j \right) \\
&= n^2 \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{\{X_i \leq x\}} (Y_i - \bar{Y}) \\
&= n^2 \left(\frac{1}{n} \sum_{i=1}^n \mathbb{1}_{\{X_i \leq x\}} Y_i - \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{\{X_i \leq x\}} \bar{Y} \right) \\
&= n^2 (H_n(x) - \bar{Y} F_n(x)) \\
&= n^2 S_n(x).
\end{aligned}$$

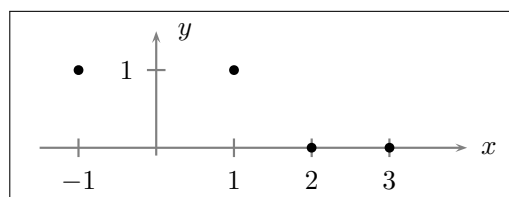
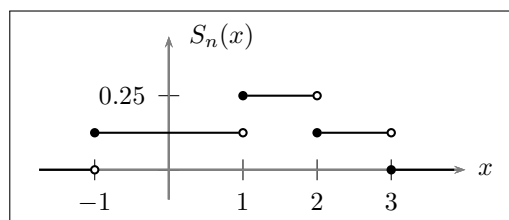
Die Multiplikation mit n^{-2} liefert die Behauptung.

Bevor wir diesen Abschnitt mit einer Betrachtung der Schreibweisen in Dempfle & Stute (2002) abschließen, wollen wir uns mit Hilfe eines Beispiels die Eigenschaften von S_n vergegenwärtigen. Darauf aufbauend liefern wir zusätzlich die Charakterisierung des argmax-Funktional nach.

Beispiel 3.8

Angenommen es gelte $a > b$, $n = 4$ und die Zufallsvariablen realisieren sich zu $(x_1, y_1) = (-1, 1)$, $(x_2, y_2) = (1, 1)$, $(x_3, y_3) = (2, 0)$ und $(x_4, y_4) = (3, 0)$. Vergleiche dazu Abbildung 3.3. Dann realisieren sich die empirischen Verteilungsfunktionen

²⁴Vgl. zu diesen Eigenschaften der Indikatorfunktion z. B. Karr (1993, S. 19).

Abbildung 3.3: Vier Realisationen des Zufallsvektors (X, Y) Abbildung 3.4: Realisierter Pfad von S_n

zu

$$F_n(x) = \begin{cases} 0 & \text{für } x < -1, \\ \frac{1}{4} & \text{für } -1 \leq x < 1, \\ \frac{1}{2} & \text{für } 1 \leq x < 2, \\ \frac{3}{4} & \text{für } 2 \leq x < 3, \\ 1 & \text{für } x \geq 3 \end{cases} \quad \text{und} \quad H_n(x) = \begin{cases} 0 & \text{für } x < -1, \\ \frac{1}{4} & \text{für } -1 \leq x < 1, \\ \frac{1}{2} & \text{für } x \geq 1 \end{cases}$$

sowie die Erfolgsquote zu $\bar{y} = \frac{1}{2}$. Damit ergibt sich für den empirischen Prozess S_n gemäß (3.14) die Realisation

$$S_n(x) = H_n(x) - \bar{y}F_n(x) = \begin{cases} 0 - \frac{1}{2} \cdot 0 = 0 & \text{für } x < -1, \\ \frac{1}{4} - \frac{1}{2} \cdot \frac{1}{4} = \frac{1}{8} & \text{für } -1 \leq x < 1, \\ \frac{1}{2} - \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4} & \text{für } 1 \leq x < 2, \\ \frac{1}{2} - \frac{1}{2} \cdot \frac{3}{4} = \frac{1}{8} & \text{für } 2 \leq x < 3, \\ \frac{1}{2} - \frac{1}{2} \cdot 1 = 0 & \text{für } x \geq 3. \end{cases}$$

In Abbildung 3.4 ist diese Funktion dargestellt.

Anhand des Beispiels werden einige Eigenschaften von S_n deutlich, die wir im Folgenden zusammentragen:²⁵

- (i) S_n ist eine zufällige Treppenfunktion, die nur in den Beobachtungen X_i springen kann. Sie ist rechtsseitig stetig mit linksseitigen Grenzwerten (rcll, *right continuous with left hand limits*).

²⁵Die Aufzählung erfolgt in Anlehnung an Ferger & Klotsche (2009, S. 98). Dort wird jedoch ein anderer empirischer Prozess betrachtet.

(ii) Außerhalb des Beobachtungsbereichs ist S_n gleich null, d. h.

$$S_n(x) = 0 \text{ für alle } x < X_{1:n} \quad \text{und} \quad S_n(x) = 0 \text{ für alle } x \geq X_{n:n}.$$

(iii) S_n besitzt keine eindeutige Maximalstelle.

(iv) S_n besitzt eine eindeutige kleinste Maximalstelle, falls der Pfad von S_n auf mindestens einem Intervall $[X_{i:n}, X_{(i+1):n}[$, $i = 1, \dots, n-1$ oberhalb der Abszisse verläuft, d. h. dort echt größer als 0 ist.²⁶

Wegen Eigenschaft (iii) ist zunächst unklar, welche Maximalstelle von S_n in (3.13) der Schätzwert für θ sein soll. Eigenschaft (iv) liefert jedoch eine Lösungsmöglichkeit.

Bemerkung 3.9 (argmax-Funktional)

Das *argmax-Funktional* gebe die kleinste Maximalstelle an, falls diese existiert. Andernfalls gebe es die kleinste Beobachtung $X_{1:n}$ an.²⁷

In den bisher verwendeten Beispielen ergeben sich damit folgende Schätzwerte:

- Beispieldatensatz A: $\hat{\theta}_n = 446$,
- Beispieldatensatz B: $\hat{\theta}_n = 499$,
- Beispiel 3.8: $\hat{\theta}_n = 1$.

Dempfle & Stute (2002) untersuchen den Schätzer $\hat{\theta}_n$ in einem allgemeineren Modellrahmen. Anstelle von (3.1) unterstellen sie lediglich das Regressionsmodell (3.6), wobei für die Zufallsvariable Y ein endlicher Erwartungswert gefordert wird. Den Spezialfall (3.1), d. h. die zusätzliche Annahme einer Bernoulli-Verteilung für Y , geben Dempfle & Stute (2002, S. 234) als illustrierendes Beispiel an. Die Regressionsfunktion κ ist auch bei Dempfle & Stute eine einstufige Treppenfunktion wie in (3.2), jedoch sind beliebige $a \neq b \in \mathbb{R}$ zugelassen.

Die Schätzung des Parameters θ basiert bei Dempfle & Stute (2002, S. 236) auf der Funktion

$$r_n(t) := n^{-2} \sum_{i=1}^n \sum_{j=1}^n (Y_i - Y_j) \mathbf{1}_{\{X_i \leq F_n^{\leftarrow}(t) < X_j\}}, \quad t \in]0, 1].$$

Im Vergleich zu S_n in der Schreibweise (3.20) hat sich die Indikatormenge geändert. Anstelle von $x \in \mathbb{R}$ werden empirische Quantile $F_n^{\leftarrow}(t) := \min\{x \in \mathbb{R} | F_n(x) \geq t\}$ mit

²⁶Dabei bezeichnet $X_{i:n}$ die i -te Ordnungsgröße. Vgl. S. 16.

²⁷Wir halten uns hier an Ferger & Klotsche (2009, S. 98). Auf eine formale Definition für alle rcll-Funktionen bzw. alle stochastischen Prozesse mit rcll-Pfaden soll hier verzichtet werden. Vgl. dazu Ferger (2009, S. 21-22) oder Ferger & Klotsche (2009, Anhang A, S. 123-124).

$t \in]0, 1]$ betrachtet.²⁸ Durch diese Transformation ist der Untersuchungsbereich stets das Einheitsintervall und damit unabhängig vom Träger von X . Für den empirischen Prozess r_n gilt

$$r_n(t) = S_n(F_n^{\leftarrow}(t)).$$

Um im Fall $a > b$ den Bruchpunkt θ zu schätzen, wird eine Maximalstelle μ_n der Funktion r_n in die empirische Quantilfunktion F_n^{\leftarrow} eingesetzt, d. h.²⁹

$$\hat{\theta}_n = F_n^{\leftarrow}(\mu_n). \quad (3.21)$$

Im weiteren Verlauf der Argumentation in Dempfle & Stute wird klar, dass μ_n keine beliebige Maximalstelle von r_n ist, sondern ein Vielfaches von $1/n$. Daher ist der Schätzwert $\hat{\theta}_n$ eine Realisation von X , denn $F_n^{\leftarrow}(i/n) = x_{i:n}$, $i = 1, \dots, n$.³⁰

Bevor die eher technische Bemerkung 3.11 nähere Informationen zur Wahl von μ_n liefert und damit den Abschnitt „Äquivalente Schreibweisen“ abschließt, soll das folgende Beispiel die gerade eingeführten Schreibweisen illustrieren.

Beispiel 3.10 (Fortsetzung von Beispiel 3.8)

In obigem Beispiel ist die empirische Quantilfunktion durch

$$F_n^{\leftarrow}(t) = \begin{cases} -1 & \text{für } 0 < t \leq \frac{1}{4}, \\ 1 & \text{für } \frac{1}{4} < t \leq \frac{1}{2}, \\ 2 & \text{für } \frac{1}{2} < t \leq \frac{3}{4}, \\ 3 & \text{für } \frac{3}{4} < t \leq 1 \end{cases}$$

gegeben. Daher realisiert sich der empirische Prozess r_n zu

$$r_n(t) = S_n(F_n^{\leftarrow}(t)) = \begin{cases} S_n(-1) = \frac{1}{8} & \text{für } 0 < t \leq \frac{1}{4}, \\ S_n(1) = \frac{1}{4} & \text{für } \frac{1}{4} < t \leq \frac{1}{2}, \\ S_n(2) = \frac{1}{8} & \text{für } \frac{1}{2} < t \leq \frac{3}{4}, \\ S_n(3) = 0 & \text{für } \frac{3}{4} < t \leq 1. \end{cases}$$

Dann ist $\mu_n = \frac{1}{2}$ die gesuchte Maximalstelle von r_n und der Schätzwert für den Parameter θ lautet

$$\hat{\theta}_n = F_n^{\leftarrow}\left(\frac{1}{2}\right) = 1.$$

²⁸Zur Quantilfunktion oder verallgemeinerten Inversen einer Verteilungsfunktion vgl. z. B. Witting (1985, S. 20, Definition 1.16).

²⁹Vgl. Dempfle & Stute (2002, S. 235, 236).

³⁰Vgl. Witting (1985, S. 28).

Bemerkung 3.11 (Maximalstelle μ_n)

Die Ausführungen von Dempfle & Stute (2002) zu μ_n bleiben in mancherlei Hinsicht vage. So lässt sich zum Beispiel nur indirekt ableiten, dass μ_n ein Vielfaches von $1/n$ sein muss:

- Dempfle & Stute definieren auf S. 236, Z. 17 die Menge

$$\Delta_n := \left\{ \frac{1}{n}, \frac{2}{n}, \dots, \frac{n-1}{n} \right\}.$$

Auf dieser Menge lässt sich der empirische Prozess r_n in geeigneter Weise anders darstellen. Damit die äquivalenten Schreibweisen ausgenutzt werden können, muss folglich auch die Maximalstelle μ_n aus der Menge Δ_n gewählt werden.

- Zusätzlich erscheint eine Erweiterung der Menge Δ_n um die Zahl 1 zu

$$\Delta_n^* := \Delta_n \cup \{1\}$$

sinnvoll. Dies erfolgt im Hinblick auf S. 238, Z. –1 wo Mengen der Form

$$\left\{ \mu_n = \frac{l}{n} \right\}$$

betrachtet werden und dabei der Fall $l = n$ explizit nicht ausgeschlossen wird. Die äquivalenten Schreibweisen von r_n und einem weiteren Prozess \tilde{r}_n auf den Seiten 236 bzw. 237 gelten nicht nur auf Δ_n , sondern auch auf der erweiterten Menge Δ_n^* .

Offen bleibt auch, was passiert, wenn r_n auf Δ_n bzw. Δ_n^* mehrere Maximalstellen besitzt. Wie lautet dann der Schätzwert für θ ?

3.2.3 Verwandte Schätzer und asymptotische Eigenschaften

In der einführenden Motivation des Dempfle-Stute-Schätzers ist bisher nur der Fall $a > b$ behandelt worden. Er tritt ein, wenn die bedingte Erfolgswahrscheinlichkeit für kleine Werte des Merkmals X größer ist als für große Werte des Merkmals X . Beispielhaft sei hier erneut das Beispiel zu Bonität und Ausfallverhalten angeführt: Bei geringer Bonität haben die Kreditnehmer eine höhere Ausfallwahrscheinlichkeit als bei höherer Bonität.

Wie jedoch ist der Parameter θ sinnvoll zu schätzen, wenn die Voraussetzung $a > b$ nicht gegeben ist und stattdessen

- a) die Annahme $a < b$ im Raum steht oder
- b) nur ein Unterschied in den Risikoniveaus, d. h. $a \neq b$, vermutet wird?

Für den Fall, dass $a < b$ gilt, deuten Dempfle & Stute (2002, S. 236) als Lösungsmöglichkeit an, das argmax-Funktional durch das argmin-Funktional zu ersetzen, d. h. die Minimalstelle von S_n als Schätzer für θ zu verwenden:

$$\hat{\theta}_n^- := \operatorname{argmin}_{x \in \mathbb{R}} S_n(x).$$

Da jede Minimalstelle von S_n eine Maximalstelle von $-S_n$ ist, lässt sich der Schätzer auch wie folgt notieren:³¹

$$\hat{\theta}_n^- = \operatorname{argmax}_{x \in \mathbb{R}} -S_n(x).$$

In Ferger (2009, S. 53) sowie in Ferger & Klotsche (2009, S. 112) wird für beliebige Konstellationen von a und b , also insbesondere für den in Buchstabe b) geschilderten Fall $a \neq b$, die zweiseitige Variante des Dempfle-Stute-Schätzers vorgeschlagen. Eine Schreibweise dieses Schätzers lautet

$$\bar{\theta}_n := \operatorname{argmax}_{x \in \mathbb{R}} |S_n(x)|.$$

Als Schätzer für θ wird also die Stelle der betragsmäßig größten Abweichung des Prozesses S_n von der x-Achse verwendet. Bemerkung 3.9 zum argmax-Funktional gilt dabei unbenommen. Eine kleinste Maximalstelle existiert nur dann nicht, wenn $S_n(x) = 0$ für alle $x \in \mathbb{R}$ gilt. Dies tritt nur in dem ungewöhnlichen Fall auf, wenn für jedes $x \in \mathbb{R}$ der Anteil der Erfolge in der Menge $\{i = 1, \dots, n | X_i \leq x\}$, d. h.

$$\frac{\sum_{i=1}^n \mathbb{1}_{\{X_i \leq x\}} Y_i}{\sum_{i=1}^n \mathbb{1}_{\{X_i \leq x\}}} = \frac{H_n(x)}{F_n(x)},$$

gleich der Gesamterfolgsquote \bar{Y} ist. Wegen Satz 3.4, S. 45 und Gleichung (2.12) auf S. 17 ist der zweiseitige Dempfle-Stute-Schätzer $\bar{\theta}_n$ die Maximalstelle der Teststatistik des Kolmogoroff-Smirnoff-(Homogenitäts-)Tests für das Hypothesenpaar in (2.10), falls $0 < \bar{Y} < 1$ gilt.

Des Weiteren stimmt die Realisation des zweiseitigen Dempfle-Stute-Schätzers stets mit der Realisation eines der beiden einseitigen Schätzer überein. Es gilt

$$\bar{\theta}_n = \begin{cases} \hat{\theta}_n, & \text{falls } \max_{x \in \mathbb{R}} S_n(x) > \max_{x \in \mathbb{R}} -S_n(x), \\ \hat{\theta}_n^-, & \text{falls } \max_{x \in \mathbb{R}} S_n(x) < \max_{x \in \mathbb{R}} -S_n(x), \\ \min\{\hat{\theta}_n, \hat{\theta}_n^-\}, & \text{falls } \max_{x \in \mathbb{R}} S_n(x) = \max_{x \in \mathbb{R}} -S_n(x). \end{cases} \quad (3.22)$$

³¹Vgl. dazu Ferger (2009, S. 59, 82) in Verbindung mit den Ausführungen in Abschnitt 3.2.2 „Äquivalente Schreibweisen“, insbesondere Gleichung (3.19).

Daraus folgt

$$\bar{\theta}_n = \hat{\theta}_n \quad \Rightarrow \quad \max_{x \in \mathbb{R}} S_n(x) \geq \max_{x \in \mathbb{R}} -S_n(x) \quad \Leftrightarrow \quad \max_{x \in \mathbb{R}} |S_n(x)| = \max_{x \in \mathbb{R}} S_n(x)$$

und

$$\bar{\theta}_n = \hat{\theta}_n^- \quad \Rightarrow \quad \max_{x \in \mathbb{R}} S_n(x) \leq \max_{x \in \mathbb{R}} -S_n(x) \quad \Leftrightarrow \quad \max_{x \in \mathbb{R}} |S_n(x)| = \max_{x \in \mathbb{R}} -S_n(x).$$

Halten wir nun einige angenehme Eigenschaften der Dempfle-Stute-Schätzer fest. Sie sind nichtparametrisch in dem Sinne, dass für die Schätzung von θ die Risikoni-
veaus a und b weder bekannt sein noch geschätzt werden müssen. Die Bestimmung des
Schätzwertes für θ ist numerisch einfach und erfolgt in einem Ein-Schritt-Verfahren. Die
Verwendung von Smoothing-Techniken, wie der Kernregression, ist dabei nicht erforder-
lich.³² Doch inwiefern sind die Schätzer gut in dem Sinne, dass sie mit wachsendem
Stichprobenumfang gegen den unbekannten Parameter konvergieren? Und wenn sie
das tun, wie schnell konvergieren sie? Diese Fragen sollen im Folgenden beantwortet
werden. Dabei zeigt sich, dass die Dempfle-Stute-Schätzer auch asymptotisch einige
wünschenswerte Eigenschaften besitzen. So sind die Schätzer unter gewissen Voraus-
setzungen stark konsistent³³ für θ (vgl. Satz 3.12). Des Weiteren erreichen sie unter
etwas stärkeren Annahmen auch die für dieses Schätzproblem optimale Konvergenzge-
schwindigkeit (vgl. Satz 3.16).

Satz 3.12 (Starke Konsistenz)

Gegeben sei das Modell (3.1) mit einstufiger Regressionsfunktion κ wie in (3.2), wobei
 $a, b \in]0, 1[$, $a \neq b$ und $\theta \in \mathbb{R}$ gelte. Zudem gelte $(X_i, Y_i) \stackrel{\text{i.i.d.}}{\sim} (X, Y)$, $i = 1, \dots, n$. Ist
die Verteilungsfunktion F für ein beliebiges $r > 0$ stetig und streng monoton wachsend
in einer Umgebung $B_r(\theta) :=]\theta - r, \theta + r[$, dann gilt

$$\begin{aligned} \hat{\theta}_n &\xrightarrow{\text{f. s.}} \theta, & \text{falls } a > b, \\ \hat{\theta}_n^- &\xrightarrow{\text{f. s.}} \theta, & \text{falls } a < b, \text{ sowie} \\ \bar{\theta}_n &\xrightarrow{\text{f. s.}} \theta, & \text{falls } a \neq b. \end{aligned}$$

Beweis: Vgl. Ferger (2009, S. 60, im Anschluss an den Beweis von Lemma 5.20).
Vgl. ebenso Ferger (2009, S. 85-86, Satz 7.3, Satz 7.6 und dessen Vorbemerkung). Dort

³²Vgl. Dempfle & Stute (2002, S. 234).

³³Eine Folge von Schätzern heißt stark konsistent, wenn sie fast sicher, d. h. mit Wahrscheinlichkeit 1,
gegen den unbekannten Parameter konvergiert. Vgl. z. B. Müller (1991, S. 318) oder Serfling (1980,
S. 48). Zur fast sicheren Konvergenz einer Folge von Zufallsvariablen vgl. z. B. Müller (1991, S. 198)
oder Serfling (1980, S. 6). Die fast sichere Konvergenz wird im Folgenden kurz als $\xrightarrow{\text{f. s.}}$ notiert. Wenn
nichts Abweichendes angegeben ist, wird dabei jeweils $n \rightarrow \infty$ unterstellt.

werden die Behauptungen in einem allgemeineren Modellrahmen bewiesen. Dieser besteht aus einem Regressionsmodell analog zu (3.6) mit weniger restriktiven Annahmen bezüglich der Regressionsfunktion κ . \square

Bemerkung 3.13 (Zur Voraussetzung von Satz 3.12)

Die Annahme des Satzes zur Verteilungsfunktion F impliziert offensichtlich $0 < F(\theta) < 1$. Damit bestätigen sich die Überlegungen zum Schätzproblem aus Abschnitt 3.1.3.

Durch Satz 3.12 wissen wir, dass die Folge der Schätzer mit Wahrscheinlichkeit 1 gegen den wahren Bruchpunkt θ konvergiert. Doch wie schnell nähert sich die Schätzerfolge an diesen wahren Parameter an? Diese Frage nach der Konvergenzgeschwindigkeit kann mit Hilfe der \mathcal{O}_P -Notation (*order in probability*) untersucht und beantwortet werden.³⁴ Diese wird in der Literatur jedoch uneinheitlich eingeführt. Die folgende Definition entspricht den Charakterisierungen in Lehmann (1999, S. 54, Definition 2.1.4(i), S. 64, Definition 2.3.1) und Bishop et al. (2007, S. 476, Definitionen 14.4-3 u. 14.4-4) unter Beachtung eines Übergangs zur Gegenwahrscheinlichkeit.³⁵

Definition 3.14 Es seien A_n und B_n , $n \in \mathbb{N}$ Folgen von reellen Zufallsvariablen auf einem gemeinsamen Wahrscheinlichkeitsraum.

- (i) Die Folge A_n heißt von der Ordnung B_n in Wahrscheinlichkeit, falls für jedes $\varepsilon > 0$ eine Konstante $\delta = \delta(\varepsilon) > 0$ und ein $n_0 = n_0(\varepsilon) \in \mathbb{N}$ existieren, so dass für schließlich alle $n > n_0$ die Aussage

$$P(|A_n| > \delta|B_n|) \leq \varepsilon$$

gilt. Dies wird kurz als $A_n = \mathcal{O}_P(B_n)$ notiert.

- (ii) Im Spezialfall $A_n = \mathcal{O}_P(1)$ heißt die Folge A_n beschränkt in Wahrscheinlichkeit oder auch stochastisch beschränkt.

Bemerkung 3.15 (\mathcal{O}_P -Notation)

- (i) Die in Definition 3.14 eingeführten Bezeichnungen lassen sich ineinander überführen. Es gilt

$$A_n = \mathcal{O}_P(B_n) \iff \frac{A_n}{B_n} = \mathcal{O}_P(1).$$

³⁴Vgl. Hamilton (1994, S. 460).

³⁵Neben der hier vorgestellten existieren weitere Charakterisierungen der \mathcal{O}_P -Notation, die sich von der hiesigen und untereinander in den Relationszeichen und teilweise in der Frage, ob die Aussage für alle oder für schließlich alle $n \in \mathbb{N}$ gelten soll, unterscheiden: Vgl. Dempfle & Stute (2002, S. 238), Brockwell & Davis (1991, S. 199, Definition 6.1.2), van der Vaart (1998, S. 8), Hamilton (1994, S. 460), White (2001, S. 28), Lehmann & Casella (1998, S. 77, Problem 8.24), Shorack (2000, S. 209, Definition 2.2(a)), Kreiß & Neuhaus (2006, S. 375, Definition A.20) oder Serfling (1980, S. 8, Abschnitt 1.2.5).

- (ii) Inhaltlich bedeutet $A_n = \mathcal{O}_P(1)$, dass die Wahrscheinlichkeitsmasse von A_n trotz anwachsendem Index n auf dem abgeschlossenen Intervall $[-\delta, \delta]$ konzentriert bleibt, also nicht zunehmend in die Verteilungsenden wandert.

Kommen wir zurück zur Frage nach der Konvergenzgeschwindigkeit. Dazu wird die Folge der Schätzfehler betrachtet. Diese ist beispielsweise für den Dempfle-Stute-Schätzer aus Definition 3.1 die Folge $\hat{\theta}_n - \theta$. Gemäß Dempfle & Stute (2002, S. 233) ist die Folge der Schätzfehler bei der Schätzung des Bruchpunkts einer Regressionsfunktion im optimalen Fall von der Ordnung n^{-1} in Wahrscheinlichkeit. Für den Schätzer $\hat{\theta}_n$ hieße das beispielhaft $\hat{\theta}_n - \theta = \mathcal{O}_P(n^{-1})$ oder gemäß Bemerkung 3.15 (i) anders ausgedrückt: $n(\hat{\theta}_n - \theta) = \mathcal{O}_P(1)$. Erfüllt ein Schätzer diese Eigenschaft heißt er auch n -konsistent.³⁶ Gemäß Bemerkung 3.15 (ii) bedeutet n -Konsistenz, dass sich der mit dem Stichprobenumfang aufgeblähte Schätzfehler, d. h. $n(\hat{\theta}_n - \theta)$, mit beliebig großer Wahrscheinlichkeit $1 - \varepsilon$ schließlich innerhalb gewisser Schranken bewegt. Der Schätzfehler muss mit wachsendem Stichprobenumfang also hinreichend schnell kleiner werden.

Der folgende Satz gibt eine Bedingung an, unter der die Dempfle-Stute-Schätzer die optimale Konvergenzgeschwindigkeit erreichen.

Satz 3.16 (Optimale Konvergenzgeschwindigkeit, n -Konsistenz)

Gegeben sei das Modell (3.1) mit einstufiger Regressionsfunktion κ wie in (3.2), wobei $a, b \in]0, 1[$, $a \neq b$ und $\theta \in \mathbb{R}$ gelte. Zudem gelte $(X_i, Y_i) \stackrel{\text{i.i.d.}}{\sim} (X, Y)$, $i = 1, \dots, n$. Ist die Verteilungsfunktion F in einer (beliebig kleinen) Umgebung von θ stetig differenzierbar mit $F'(\theta) > 0$, dann gilt

$$n(\hat{\theta}_n - \theta) = \mathcal{O}_P(1), \quad \text{falls } a > b, \quad (3.23)$$

$$n(\hat{\theta}_n^- - \theta) = \mathcal{O}_P(1), \quad \text{falls } a < b, \text{ sowie} \quad (3.24)$$

$$n(\bar{\theta}_n - \theta) = \mathcal{O}_P(1), \quad \text{falls } a \neq b. \quad (3.25)$$

Beweis: Zum Beweis von (3.23) und (3.24) vgl. das Theorem in Dempfle & Stute (2002, S. 237). Das hier gegebene Modell ist ein Spezialfall des dort behandelten Modellrahmens. Gegenüber diesem wird hier gemäß (3.1) zusätzlich eine Bernoulli-Verteilung für die Zufallsvariable Y angenommen. Dies führt automatisch zur Erfüllung der ersten Annahme (A1) des Theorems, die besagt, dass die bedingte Varianz von Y gegeben $X = x$ beschränkt sein muss, denn gemäß (3.7) gilt

$$\mathbb{V}[Y|X = x] = \kappa(x)(1 - \kappa(x)) \leq \max\{a(1 - a), b(1 - b)\} \leq \frac{1}{4} < \infty.$$

³⁶Vgl. Ferger & Klotsche (2009, S. 101). Siehe auch Lehmann (1999, S. 475) zum analog verwendeten Begriff \sqrt{n} -Konsistenz.

Daher muss nur noch die zweite Bedingung (A2) im Theorem von Dempfle & Stute geprüft werden. Diese stimmt jedoch mit der im Satz 3.16 formulierten Voraussetzung bezüglich der Verteilungsfunktion F überein. Damit sind (3.23) und (3.24) gezeigt.

Bevor wir zum Beweis von (3.25) kommen, folgt zunächst eine Bemerkung zum Beweis in Dempfle & Stute.

Bemerkung 3.17 (Zum Beweis in Dempfle & Stute (2002))

- (i) Im Beweis von Dempfle & Stute (2002, S. 238-241) wird ausführlich nur der Fall $a > b$ behandelt. Der Fall $a < b$ wird analog bewiesen und zwar durch Übergang zum $\arg\min$ -Funktional wie in den Präliminarien angedeutet. Möglich wäre auch die Spiegelung der zugrunde liegenden (empirischen) Prozesse an der Abszisse und deren Maximierung.
- (ii) Der Beweis von Dempfle & Stute (2002, S. 238-241) enthält eine (behebbar) Lücke: Um die stochastische Beschränktheit von $n(\hat{\theta}_n - \theta)$ zu zeigen, betrachten Dempfle & Stute die Menge $\{n|\hat{\theta}_n - \theta| \geq T\}$. Auf S. 239, Z. 9, führen sie dann die Funktion γ ein. Es ist

$$\gamma(l) = (a - b)F_n(\theta)(l - nF_n(\theta)).$$

Auf Z. 10 wird anschließend festgestellt:

„Since by assumption $a > b$ we have $\gamma(l) > 0$ on the set $\{F_n(\theta) > 0\}$.“

Diese Aussage ist falsch, denn es ist nicht sichergestellt, dass der Term $l - nF_n(\theta)$ ungleich null ist. Bekannt ist nur $l \geq nF_n(\theta)$, benötigt wird aber die strikte Ungleichheit.

Diese Lücke soll nun behoben werden. Für den Nachweis stochastischer Beschränktheit genügt es, die Menge $\{n|\hat{\theta}_n - \theta| > T\}$ zu betrachten, d. h. $>$ anstelle von \geq zu verwenden.³⁷ Für das in Dempfle & Stute betrachtete obere Verteilungsende gilt dann

$$\begin{aligned} \{n(\hat{\theta}_n - \theta) > T\} &= \{F_n^{\leftarrow}(\mu_n) > \theta + \frac{T}{n}\} \\ &= \{\mu_n > F_n(\theta + \frac{T}{n})\} \\ &= \bigcup_{nF_n(\theta + \frac{T}{n}) < l \leq n} \{\mu_n = \frac{l}{n}\}. \end{aligned}$$

³⁷Vgl. Definition 3.14.

Dabei gilt das erste Gleichheitszeichen wegen (3.21). Das zweite folgt aus Aussage a) in Hilfssatz 1.17 in Witting (1985, S. 20), denn deren Negation lautet

$$F^{\leftarrow}(y) > x \Leftrightarrow y > F(x)$$

für alle $y \in]0, 1]$, für alle $x \in \mathbb{R}$ und für jede Verteilungsfunktion F . Das letzte Gleichheitszeichen gilt, weil μ_n ein Vielfaches von $\frac{1}{n}$ ist.³⁸ Aus $l > nF_n(\theta + \frac{T}{n}) \geq nF_n(\theta)$ folgt wie gewünscht $l > nF_n(\theta)$. Damit ist die Lücke im Beweis geschlossen.

Die noch offene Aussage (3.25) folgt aus Theorem 8.9 (1) in Ferger (2009, S. 107) mittels des Grenzübergangs $x \rightarrow \infty$.³⁹ Das Theorem setzt einen allgemeineren Modellrahmen als wir voraus. Dieser besteht aus einem Regressionsmodell analog zu (3.6) mit weniger restriktiven Annahmen bezüglich der Regressionsfunktion κ .⁴⁰ Auch die Voraussetzungen an die Verteilungsfunktion F sind allgemeiner gehalten. Es wird lediglich gefordert,⁴¹

- dass F in einer gelochten Umgebung $B_r(\theta) \setminus \{\theta\}$ von θ differenzierbar ist,
- dass die links- und die rechtsseitige Ableitung an der Stelle θ positiv sind, d. h. $F'(\theta \pm) > 0$ gilt, und
- dass die Ableitung F' auf $B_r(\theta) \setminus \{\theta\}$ beschränkt ist.

Diese drei Punkte sind aufgrund unserer Annahmen an F , insbesondere der stetigen Differenzierbarkeit, erfüllt. Damit ist auch (3.25) gezeigt und der Satz vollständig bewiesen. \square

Bemerkung 3.18 (Zu den Voraussetzungen von Satz 3.16)

- (i) Die letzten Ausführungen des Beweises zeigen, dass die Voraussetzungen des Satzes auch hätten allgemeiner gehalten werden können. Dies wurde aus Gründen der Übersichtlichkeit und Einfachheit unterlassen.
- (ii) Die Voraussetzungen von Satz 3.16 implizieren die Voraussetzungen von Satz 3.12 zur starken Konsistenz, da Differenzierbarkeit Stetigkeit impliziert.⁴²

³⁸Vgl. dazu auch Bemerkung 3.11.

³⁹Vgl. dazu auch den Übergang von (5.5) zu (5.6) in Ferger (2009, S. 40).

⁴⁰Vgl. dazu Ferger (2009, S. 82-87).

⁴¹Diese Voraussetzungen schließen im Gegensatz zu unserer Annahme an F z. B. nicht aus, dass F an der Stelle θ unstetig ist.

⁴²Vgl. z. B. Heuser (1994, S. 262, Satz 46.1).

3.3 Maximum-Likelihood-Schätzer bei bekannten Risikoniveaus

Ferger & Klotsche (2009, S. 97-111) untersuchen in ihrer Arbeit⁴³ den Maximum-Likelihood-Schätzer im Modell (3.1) mit κ gemäß (3.2). Dabei unterstellen sie $(X_i, Y_i) \stackrel{\text{i.i.d.}}{\sim} (X, Y)$, $i = 1, \dots, n$. Des Weiteren werden die Risikoniveaus a und b als bekannt vorausgesetzt. Als Likelihoodfunktion ergibt sich⁴⁴

$$\mathcal{L}_n(\theta) = \prod_{i=1}^n \kappa(x_i)^{y_i} (1 - \kappa(x_i))^{1-y_i}.$$

Die Loglikelihoodfunktion

$$\ell_n(\theta) = \ln \mathcal{L}_n(\theta) = \sum_{i=1}^n (y_i \ln \kappa(x_i) + (1 - y_i) \ln(1 - \kappa(x_i)))$$

lässt sich wegen der speziellen Treppenform von κ umschreiben zu

$$\ell_n(\theta) = \sum_{i=1}^n \mathbb{1}_{\{x_i \leq \theta\}} \left(y_i \ln \left(\frac{a}{b} \frac{1-b}{1-a} \right) + \ln \frac{1-a}{1-b} \right) + \sum_{i=1}^n (y_i \ln b + (1 - y_i) \ln(1 - b)). \quad (3.26)$$

Bemerkung 3.19 (Zum Beweis von (3.26))

Vergleiche zu einigen Zwischenergebnissen der folgenden Rechnung auch Ferger & Klotsche (2009, S. 98) sowie Ferger (2009, S. 35). Es gilt

$$\begin{aligned} \ell_n(\theta) &= \sum_{i=1}^n (y_i \ln \kappa(x_i) + (1 - y_i) \ln(1 - \kappa(x_i))) \\ &= \sum_{i=1}^n (\mathbb{1}_{\{x_i \leq \theta\}} + \mathbb{1}_{\{x_i > \theta\}}) (y_i \ln \kappa(x_i) + (1 - y_i) \ln(1 - \kappa(x_i))) \\ &= \sum_{i=1}^n \mathbb{1}_{\{x_i \leq \theta\}} (y_i \ln \kappa(x_i) + (1 - y_i) \ln(1 - \kappa(x_i))) \\ &\quad + \sum_{i=1}^n \mathbb{1}_{\{x_i > \theta\}} (y_i \ln \kappa(x_i) + (1 - y_i) \ln(1 - \kappa(x_i))) \end{aligned}$$

⁴³Diese Arbeit basiert auf Ferger (2009, S. 34-50). Für Quellenangaben wird deshalb der Einfachheit halber nur auf Ferger & Klotsche (2009) verwiesen, es sei denn, in Ferger (2009) befinden sich ausführlichere Informationen.

⁴⁴Vgl. Ferger & Klotsche (2009, S. 97) sowie Ferger (2009, S. 34).

$$\begin{aligned}
&= \sum_{i=1}^n \mathbb{1}_{\{x_i \leq \theta\}} (y_i \ln a + (1 - y_i) \ln(1 - a)) \\
&\quad + \sum_{i=1}^n \mathbb{1}_{\{x_i > \theta\}} (y_i \ln b + (1 - y_i) \ln(1 - b)) \\
&= \sum_{i=1}^n \mathbb{1}_{\{x_i \leq \theta\}} (y_i \ln a + (1 - y_i) \ln(1 - a)) \\
&\quad + \sum_{i=1}^n (1 - \mathbb{1}_{\{x_i \leq \theta\}}) (y_i \ln b + (1 - y_i) \ln(1 - b)) \\
&= \sum_{i=1}^n \mathbb{1}_{\{x_i \leq \theta\}} (y_i \ln a + (1 - y_i) \ln(1 - a) - y_i \ln b - (1 - y_i) \ln(1 - b)) \\
&\quad + \sum_{i=1}^n (y_i \ln b + (1 - y_i) \ln(1 - b)) \\
&= \sum_{i=1}^n \mathbb{1}_{\{x_i \leq \theta\}} \left(y_i \ln \frac{a}{b} + (1 - y_i) \ln \frac{1 - a}{1 - b} \right) + \sum_{i=1}^n (y_i \ln b + (1 - y_i) \ln(1 - b)) \\
&= \sum_{i=1}^n \mathbb{1}_{\{x_i \leq \theta\}} \left(y_i \ln \left(\frac{a}{b} \frac{1 - b}{1 - a} \right) + \ln \frac{1 - a}{1 - b} \right) + \sum_{i=1}^n (y_i \ln b + (1 - y_i) \ln(1 - b)).
\end{aligned}$$

Damit ist (3.26) gezeigt.

Der zweite Teil der Loglikelihoodfunktion in (3.26) enthält den Parameter θ nicht. Daher genügt es, den ersten Teil der Summe zu maximieren, um den **Maximum-Likelihood-Schätzer** (ML-Schätzer) für θ zu erhalten: Definiere⁴⁵

$$\theta_n^* := \operatorname{argmax}_{x \in \mathbb{R}} S_n^*(x),$$

wobei

$$S_n^*(x) := \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{\{X_i \leq x\}} (\alpha Y_i + \beta) \quad (3.27)$$

mit

$$\alpha := \ln \left(\frac{a}{b} \frac{1 - b}{1 - a} \right) \quad \text{und} \quad \beta := \ln \frac{1 - a}{1 - b}. \quad (3.28)$$

Bemerkung 3.20 (Zur Definition des ML-Schätzers)

- (i) Die Bemerkung 3.9 zum *argmax*-Funktional gilt hier unverändert.
- (ii) Die Realisation der zufälligen Funktion S_n^* unterscheidet sich vom ersten Teil der Loglikelihoodfunktion durch den Faktor $\frac{1}{n}$. Dieser hat keinen Einfluss auf die

⁴⁵Vgl. Ferger & Klotsche (2009, S. 98).

Maximalstelle der betrachteten Funktionen. Damit ist die Bezeichnung von θ_n^* als ML-Schätzer gerechtfertigt.

- (iii) Mit den zufälligen empirischen Verteilungsfunktionen F_n und H_n gemäß (2.1) bzw. (3.12) ergibt sich für den empirischen Prozess S_n^* die alternative Schreibweise

$$S_n^*(x) = \alpha H_n(x) + \beta F_n(x), \quad x \in \mathbb{R}.$$

- (iv) Wegen

$$a \neq b \iff 1 - a \neq 1 - b \iff a - ab \neq b - ab$$

folgt

$$\alpha \neq 0, \quad \beta \neq 0 \quad \text{und} \quad \alpha + \beta = \ln \frac{a}{b} \neq 0. \quad (3.29)$$

Insbesondere aufgrund des vorletzten Punktes der Bemerkung lassen sich folgende Eigenschaften des empirischen Prozesses S_n^* ableiten:⁴⁶

- (i) S_n^* ist eine zufällige rcll-Treppenfunktion, die nur in den Beobachtungen X_i springen kann.
- (ii) Es gilt $S_n^*(x) = 0$ für alle $x < X_{1:n}$ und $S_n^*(x) = \alpha \bar{Y} + \beta$ für alle $x \geq X_{n:n}$.
- (iii) S_n^* besitzt keine eindeutige Maximalstelle.
- (iv) S_n^* besitzt eine eindeutige kleinste Maximalstelle, falls der Pfad von S_n^* auf mindestens einem Intervall $[X_{i:n}, X_{(i+1):n}]$, $i = 1, \dots, n-1$, oder $[X_{n:n}, +\infty[$ oberhalb der Abszisse (x-Achse) verläuft, d. h. dort echt größer als 0 ist.

Im Folgenden werden die asymptotischen Eigenschaften des ML-Schätzers θ_n^* untersucht. Dabei wird nachstehende Annahme über die Verteilungsfunktion F genutzt.⁴⁷

Annahme 3.21

Es existiert kein Intervall, welches θ enthält und auf dem F konstant ist.

Diese Annahme impliziert $0 < F(\theta) < 1$ und somit die schon mehrfach angesprochene Mindestvoraussetzung unseres Schätzproblems. Die Voraussetzungen der Sätze 3.12 und 3.16 implizieren Annahme 3.21. Im Gegensatz zu ihnen fordert Annahme 3.21 nicht, dass F im Punkt θ stetig ist. Ist die Annahme nicht erfüllt, weil der unbekannte

⁴⁶Vgl. dazu auch Ferger & Klotsche (2009, S. 98).

⁴⁷Vgl. hier und im Folgenden Ferger & Klotsche (2009, S. 99).

Bruchpunkt θ im Inneren eines Intervalls liegt, auf dem F konstant ist, dann gibt es eine Umgebung von θ , welche fast sicher keine Beobachtungsvariable X_i enthält.⁴⁸ Somit wäre es unmöglich, die exakte Lage von θ innerhalb dieser Umgebung zu schätzen. Die Güte der Schätzung wäre daher an die Grenzen des Intervalls gebunden.

Der folgende Satz über die starke Konsistenz des ML-Schätzers verwendet Annahme 3.21.

Satz 3.22 (Starke Konsistenz)

Gegeben sei das Modell (3.1) mit einstufiger Regressionsfunktion κ wie in (3.2), wobei $a, b \in]0, 1[$, $a \neq b$ und $\theta \in \mathbb{R}$ gelte. Zudem gelte $(X_i, Y_i) \stackrel{\text{i.i.d.}}{\sim} (X, Y)$, $i = 1, \dots, n$. Erfüllt die Verteilungsfunktion F die Annahme 3.21, dann gilt

$$\theta_n^* \xrightarrow{\text{f. s.}} \theta.$$

Beweis: Vgl. Ferger & Klotsche (2009, S. 100, Proposition 2.3). □

Eine hinreichende Bedingung für die n -Konsistenz des ML-Schätzers wird im nachstehenden Satz angegeben.

Satz 3.23 (Konvergenzgeschwindigkeit, n -Konsistenz)

Gegeben sei das Modell (3.1) mit einstufiger Regressionsfunktion κ wie in (3.2), wobei $a, b \in]0, 1[$, $a \neq b$ und $\theta \in \mathbb{R}$ gelte. Zudem gelte $(X_i, Y_i) \stackrel{\text{i.i.d.}}{\sim} (X, Y)$, $i = 1, \dots, n$. Falls es ein $r > 0$ und positive Konstanten $\underline{L} = \underline{L}(r)$ und $\bar{L} = \bar{L}(r)$ derart gibt, dass

$$\underline{L}|s - t| \leq |F(s) - F(t)| \leq \bar{L}|s - t| \text{ für alle } s, t \in B_r(\theta) \quad (3.30)$$

gilt, dann gilt

$$n(\theta_n^* - \theta) = \mathcal{O}_P(1).$$

Beweis: Vgl. Ferger & Klotsche (2009, S. 101, Proposition 3.1 und Corollary 3.2) in Verbindung mit Bemerkung 3.15(i). Im Beweis von Proposition 3.1 befinden sich mehrere kleine Schreib-/Setzfehler, die hier dokumentiert werden sollen:⁴⁹

- S. 101, Gleichung (3.4): $a_n := P(|\hat{\theta}_n - \theta| > r)$, vgl. dazu die letzten vier Zeilen des Beweises auf S. 105,
- S. 102, Z. 10: $S = \alpha H + \beta F$,
- S. 104, Z. 6 und -3: $1 \leq k \leq 2^m$,

⁴⁸In Ferger & Klotsche (2009, S. 99) heißt es: Ist die Annahme nicht erfüllt, dann liegt der unbekannte Bruchpunkt θ in einem Intervall, welches fast sicher keine Beobachtungsvariable X_i enthält. Diese Aussage ist dann nicht richtig, wenn auf θ Wahrscheinlichkeitsmasse liegt.

⁴⁹Vgl. hierzu auch Ferger (2009, S. 40-44, Beweis von Satz 5.5).

- S. 104, Z. –8: $k = 1$.

Zudem ist die Bezeichnung der Konstante A in den letzten vier Zeilen des Beweises unglücklich. Die Konstante ist nicht zu verwechseln mit der Wahrscheinlichkeit A , die auf S. 105, Ungleichung (3.5) definiert wurde. \square

Bemerkung 3.24 (Zur Voraussetzung von Satz 3.23)

- (i) Die zweite Ungleichung in (3.30) ist die sogenannte *Dehnungsbeschränktheit* oder *Lipschitz-Stetigkeit* von F auf der Menge $B_r(\theta)$. Sie impliziert die Stetigkeit von F auf $B_r(\theta)$.⁵⁰
- (ii) In (3.30) gelte ohne Beschränkung der Allgemeinheit $s > t$. Dann lässt sich (3.30) wegen der Monotonie der Verteilungsfunktion F umschreiben zu:

$$0 < \underline{L} \leq \frac{F(s) - F(t)}{s - t} \leq \bar{L} \text{ für alle } s > t \in B_r(\theta). \quad (3.31)$$

Der Quotient in (3.31) ist der Anstieg der Sekante durch die Punkte $(t, F(t))$ und $(s, F(s))$. Dieser Anstieg ist für jede beliebige Wahl von $s > t$ echt positiv. Daher ist die Verteilungsfunktion auf $B_r(\theta)$ streng monoton wachsend. Somit folgt:

- (iii) Die Voraussetzungen von Satz 3.23 implizieren Annahme 3.21 und damit die Voraussetzung von Satz 3.22 zur starken Konsistenz des ML-Schätzers.

Es folgt ein Ergebnis zur asymptotischen Verteilung der Folge $n(\theta_n^* - \theta)$. Zum Begriff der fast schwachen Konvergenz vergleiche anschließend Bemerkung 3.27.

Satz 3.25 (Fast schwache Konvergenz)

Gegeben sei das Modell (3.1) mit einstufiger Regressionsfunktion κ wie in (3.2), wobei $a, b \in]0, 1[$, $a \neq b$ und $\theta \in \mathbb{R}$ gelte. Zudem gelte $(X_i, Y_i) \stackrel{\text{i.i.d.}}{\sim} (X, Y)$, $i = 1, \dots, n$. Falls die Verteilungsfunktion F die Bedingung (3.30) erfüllt und an der Stelle θ sowohl eine rechtsseitige Ableitung $F'(\theta+)$ als auch eine linksseitige Ableitung $F'(\theta-)$ besitzt, dann gelten

$$\limsup_{n \rightarrow \infty} P(n(\theta_n^* - \theta) \leq x) \leq P(\tau \leq x) \quad (3.32)$$

und

$$\liminf_{n \rightarrow \infty} P(n(\theta_n^* - \theta) < x) \geq P(\sigma < x) \quad (3.33)$$

⁵⁰Vgl. z. B. Heuser (1994, S. 136, 212).

für alle $x \in \mathbb{R}$. Dabei bezeichnen τ und σ die kleinste bzw. größte Maximalstelle des zusammengesetzten Poisson-Prozesses⁵¹ L mit

$$L(x) = \begin{cases} \sum_{i=1}^{N_1(x)} \xi_i, & \text{falls } x \geq 0, \\ -\sum_{i=1}^{N_2(-x)} \eta_i, & \text{falls } x < 0. \end{cases} \quad (3.34)$$

Hierbei

- ist N_1 ein rcll-Poisson-Prozess⁵² mit Parameter $\lambda_1 = F'(\theta+)$,
- ist N_2 ein lcr⁵³-Poisson-Prozess mit Parameter $\lambda_2 = F'(\theta-)$,
- besitzen die ξ_i für jedes $i \in \mathbb{N}$ eine Zweipunktverteilung mit

$$P(\xi_i = \alpha + \beta) = b, \quad P(\xi_i = \beta) = 1 - b, \quad (3.35)$$

- besitzen die η_i für jedes $i \in \mathbb{N}$ eine Zweipunktverteilung mit

$$P(\eta_i = \alpha + \beta) = a, \quad P(\eta_i = \beta) = 1 - a, \quad (3.36)$$

- sind die beteiligten Zufallsgrößen $N_1, N_2, \xi_i, \eta_i, i \in \mathbb{N}$ insgesamt stochastisch unabhängig.

Beweis: Vgl. Ferger & Klotsche (2009, S. 105-111, insb. Theorem 4.4). Auch hierzu gibt es einige Anmerkungen:

- Auf S. 109, Z. 4 muss es korrekterweise

$$(L_n(s), L_n(t) - L_n(s)) \xrightarrow{\mathcal{L}} (L(s), L(t) - L(s))$$

heißen.⁵⁴

⁵¹Zum Begriff des zusammengesetzten Poisson-Prozesses vgl. z. B. Meintrup & Schöffler (2005, S. 287).

⁵²Vgl. z. B. Meintrup & Schöffler (2005, Kap. 10, S. 267-279) oder Müller (1991, S. 298-300).

⁵³Die Abkürzung lcr⁵³ steht für left continuous with right-hand limits, zu deutsch: linksseitig stetig mit rechtsseitigen Grenzwerten. Diese Forderung führt dazu, dass L als Gesamtheit rcll ist.

⁵⁴Vgl. auch Ferger (2009, S. 48). Dabei bezeichnet $\xrightarrow{\mathcal{L}}$ die Konvergenz in Verteilung. Diese Konvergenzart (vgl. z. B. Karr (1993, S. 136, Definition 5.5 und S. 149, Definition 5.26) oder Dudewicz & Mishra (1988, S. 300, Definition 6.24)) wird im Folgenden durch \xrightarrow{d} bezeichnet. Wenn nichts Abweichendes angegeben ist, wird dabei jeweils $n \rightarrow \infty$ unterstellt.

- In Lemma 4.2 (S. 110) ist die Konstante bei der Abschätzung unvollständig. Richtigerweise muss es

$$\mathbb{E}[|L_n(t) - L_n(t_1)| |L_n(t_2) - L_n(t)|] \leq c^2 \bar{L}^2(t - t_1)(t_2 - t)$$

heißen, wobei $c := |\alpha + \beta| + |\beta|$. Wegen (3.29) ist $c > 0$.

Dieser Fehler begründet sich darin, dass im Beweis von Lemma 4.2 das erste Gleichheitszeichen falsch ist. Statt Y_i müsste es dort wegen S. 107, Z. 2 korrekterweise $(\alpha Y_i + \beta)$ heißen. Wegen $Y_i \in \{0, 1\}$ kann der Ausdruck $|\alpha Y_i + \beta|$ durch c nach oben abgeschätzt werden. Alsdann wird der Faktor c aus der Summe und dem Erwartungswert herausgezogen. Analoges gilt für Y_j . Damit ergibt sich der zusätzliche Faktor c^2 für die Aussage von Lemma 4.2.⁵⁵

□

Bemerkung 3.26 (Zu den Voraussetzungen von Satz 3.25)

Die Annahme von Bedingung (3.30) stellt sicher, dass sowohl die rechts- als auch die linksseitige Ableitung von F an der Stelle θ positiv und endlich ist, sofern sie existiert.⁵⁶ Die zusätzliche Forderung der Existenz der rechts- und linksseitigen Ableitung macht die Voraussetzungen von Satz 3.25 schärfer als die von Satz 3.23 zur n -Konsistenz des ML-Schätzers. Gemäß Bemerkung 3.24(iii) sind die Voraussetzungen dann auch schärfer als die von Satz 3.22 zur starken Konsistenz des ML-Schätzers.

Bemerkung 3.27 (Fast schwache Konvergenz)

Die Eigenschaften in (3.32) und (3.33) werden als **fast schwache Konvergenz** (almost weak convergence) von $n(\theta_n^* - \theta)$ gegen das zufällige Paar (τ, σ) bezeichnet und als

$$n(\theta_n^* - \theta) \rightsquigarrow_{a.w.} (\tau, \sigma)$$

notiert.⁵⁷

Als Folge der fast schwachen Konvergenz gemäß (3.32) und (3.33) lässt sich ein asymptotisches Konfidenzintervall für θ angeben.

Folgerung 3.28 (Asymptotisches Konfidenzintervall für θ)

Gegeben seien die Voraussetzungen und Bezeichnungen von Satz 3.25. Für das Paar (c_1, d_1) gelte $P(\sigma < c_1) - P(\tau \leq -d_1) \geq 1 - \gamma$, wobei $0 < \gamma < 1$. Dann ist das Intervall

$$I_n(\gamma) = \left] \theta_n^* - \frac{c_1}{n}, \theta_n^* + \frac{d_1}{n} \right[\quad (3.37)$$

⁵⁵Vgl. auch Ferger (2009, S. 48-49). Dort fehlen jedoch in den Abschätzungen in der vorletzten Zeile des Beweises die Betragszeichen in den Zählern.

⁵⁶Vgl. Ferger & Klotsche (2009, S. 106).

⁵⁷Vgl. Ferger & Klotsche (2009, S. 95, Definition 1.1 und S. 111, Remark 4.5(1)).

ein asymptotisches Konfidenzintervall für θ zum Konfidenzniveau $1 - \gamma$, d. h.

$$\liminf_{n \rightarrow \infty} P(\theta \in I_n(\gamma)) \geq 1 - \gamma.$$

Beweis: Vgl. Ferger & Klotsche (2009, S. 111, Remark 4.5(2)) oder, etwas ausführlicher, Ferger (2009, S. 50, Bemerkung 5.10(2)). In beiden Arbeiten wird das Konfidenzniveau unglücklicherweise mit $1 - \alpha$ bezeichnet, obwohl α auch dort für die Größe aus (3.28) verwendet wird. In Ferger (2009) finden sich zudem die Bezeichnungen a und b in den Grenzen des Konfidenzintervalls. Diese Notationen sind nicht zu verwechseln mit den Risikoniveaus a und b . \square

Zur Bestimmung der Zahlen c_1 und d_1 benötigt man Informationen über die Verteilung der Zufallsvariablen τ und σ , welche bisher nur implizit als die kleinste bzw. größte Maximalstelle des zusammengesetzten Poisson-Prozesses L aus (3.34) eingeführt wurden. Ferger & Klotsche (2009, S. 117-118) geben explizite Darstellungen der Zufallsvariablen τ und σ an, mit deren Hilfe eine Monte-Carlo-Approximation für die Zahlen c_1 und d_1 durchgeführt werden kann.

Bemerkung 3.29 (Explizite Darstellungen der Zufallsvariablen τ und σ)

Auf Grundlage der Zufallsvariablen ξ_i und η_i , $i \in \mathbb{N}$ aus Satz 3.25 werden zunächst die Partialsummen

$$V_k := \sum_{i=1}^k \xi_i, \quad k \in \mathbb{N}_0 \quad \text{und} \quad W_m := - \sum_{i=1}^m \eta_i, \quad m \in \mathbb{N}_0 \quad (3.38)$$

definiert. Weil die beiden Trägerpunkte der Zufallsvariablen ξ_i und η_i , $i \in \mathbb{N}$, d. h. $\alpha + \beta = \ln(\frac{a}{b}) \neq 0$ und $\beta = \ln(\frac{1-a}{1-b}) \neq 0$, verschieden von null sind, verändern die Partialsummen aus (3.38) in jedem $k \in \mathbb{N}$ bzw. in jedem $m \in \mathbb{N}$ ihren Wert. Wegen (3.35) und (3.36) in dieser Arbeit sowie (2.7) in Ferger & Klotsche (2009, S. 99) gelten

$$\mathbb{E}[\xi_i] = \mathbb{E}[\xi_1] = (\alpha + \beta)b + \beta(1 - b) = \alpha b + \beta < 0$$

und

$$\mathbb{E}[\eta_i] = \mathbb{E}[\eta_1] = (\alpha + \beta)a + \beta(1 - a) = \alpha a + \beta > 0$$

für alle $i \in \mathbb{N}$. Mit dem starken Gesetz der Großen Zahlen folgt daraus, dass für die Partialsummen

$$V_k \xrightarrow{\text{f. s.}} -\infty \quad (k \rightarrow \infty) \quad \text{und} \quad W_m \xrightarrow{\text{f. s.}} -\infty \quad (m \rightarrow \infty)$$

gilt. Somit existieren die Zufallsvariablen

$$V_{\max} := \max_{k \geq 0} V_k \quad \text{und} \quad W_{\max} := \max_{m \geq 0} W_m$$

fast sicher und sind zudem fast sicher endlich. Daher existieren fast sicher auch der kleinste und größte Index, bei denen die Maxima V_{\max} und W_{\max} angenommen werden. Diese Indizes werden wie folgt bezeichnet:

$$\begin{aligned} \underline{k} &:= \min\{k \in \mathbb{N}_0 | V_k = V_{\max}\}, \\ \bar{k} &:= \max\{k \in \mathbb{N}_0 | V_k = V_{\max}\}, \\ \underline{m} &:= \min\{m \in \mathbb{N}_0 | W_m = W_{\max}\}, \\ \bar{m} &:= \max\{m \in \mathbb{N}_0 | W_m = W_{\max}\}. \end{aligned}$$

Des Weiteren bezeichnen $0 =: T_0 < T_1 < T_2 < \dots$ und $0 =: U_0 < U_1 < U_2 < \dots$ die zu den Poisson-Prozessen N_1 bzw. N_2 gehörigen Sprungzeiten (Ereigniszeitpunkte).⁵⁸ Somit ergeben sich für die kleinste bzw. größte Maximalstelle des zusammengesetzten Poisson-Prozesses L aus (3.34) die Gleichungen

$$\tau = \begin{cases} T_{\underline{k}}, & \text{falls } V_{\max} > W_{\max}, \\ -U_{\bar{m}+1}, & \text{falls } V_{\max} \leq W_{\max}, \end{cases} \quad \text{und} \quad \sigma = \begin{cases} T_{\bar{k}+1}, & \text{falls } V_{\max} \geq W_{\max}, \\ -U_{\underline{m}}, & \text{falls } V_{\max} < W_{\max}. \end{cases}$$

Zur Simulation der Verteilung der Zufallsvariablen τ und σ werden die Parameter a und b sowie $\lambda_1 = F'(\theta+)$ und $\lambda_2 = F'(\theta-)$ benötigt. Die Risikoniveaus a und b sind bekannt. Die Intensitäten der Poisson-Prozesse λ_1 und λ_2 sind jedoch unbekannt und müssen daher durch Schätzwerte ersetzt werden. Dazu wird der Einfachheit halber angenommen, dass die Zufallsvariable X die stetige Dichte f besitze, so dass $\lambda_1 = \lambda_2 = f(\theta)$ gilt. Ist f_n ein Dichteschätzer, der auf einer Umgebung $B_r(\theta)$, $r > 0$ fast sicher gleichmäßig gegen f konvergiert, dann führt $\theta_n^* \xrightarrow{\text{f. s.}} \theta$ zu

$$\lambda_n^* := f_n(\theta_n^*) \xrightarrow{\text{f. s.}} f(\theta). \quad (3.39)$$

Folglich können λ_1 und λ_2 jeweils durch λ_n^* ersetzt werden.⁵⁹

⁵⁸Die Zwischenereigniszeiten (interarrival times, waiting times) $T_{k+1} - T_k$, $k \in \mathbb{N}_0$ bzw. $U_{m+1} - U_m$, $m \in \mathbb{N}_0$ sind exponentialverteilt. Die Parameter der Exponentialverteilungen sind gerade die Intensitäten der Poisson-Prozesse λ_1 bzw. λ_2 .

⁵⁹Vgl. dazu Ferger & Klotzsche (2009, S. 118). Dort ist das Vorgehen im Rahmen des hier in Abschnitt 3.4 behandelten Substitutionsschätzers beschrieben.

Bemerkung 3.30 (Zum Beweis von (3.39))

Gegeben seien eine stetige Dichte f , ein Dichteschätzer f_n , welcher auf einer Umgebung $B_r(\theta)$, $r > 0$ fast sicher gleichmäßig gegen f konvergiert, d. h.

$$\sup_{x \in B_r(\theta)} |f_n(x) - f(x)| \xrightarrow{\text{f. s.}} 0, \quad (3.40)$$

und die starke Konsistenz des ML-Schätzers:

$$\theta_n^* \xrightarrow{\text{f. s.}} \theta. \quad (3.41)$$

Zum Beweis von (3.39) verwenden wir folgende Abschätzung:⁶⁰

$$\begin{aligned} 0 \leq |f_n(\theta_n^*) - f(\theta)| &= |f_n(\theta_n^*) - f(\theta_n^*) + f(\theta_n^*) - f(\theta)| \\ &\leq \underbrace{|f_n(\theta_n^*) - f(\theta_n^*)|}_{\xrightarrow{\text{f. s.}} 0} + \underbrace{|f(\theta_n^*) - f(\theta)|}_{\xrightarrow{\text{f. s.}} 0}. \end{aligned}$$

Wegen (3.41) gilt mit Wahrscheinlichkeit eins (fast sicher) die Beziehung $\theta_n^* \in B_r(\theta)$ für schließlich alle $n \in \mathbb{N}$. Daher konvergiert der erste Summand der Oberschranke fast sicher gegen null gemäß (3.40). Der zweite Summand konvergiert fast sicher gegen null, weil f stetig ist und somit die Bedingungen des sogenannten Continuous Mapping Theorems (Stetigkeitssatzes)⁶¹ erfüllt sind. Damit folgt aus der Konvergenz in (3.41) die Aussage $f(\theta_n^*) \xrightarrow{\text{f. s.}} f(\theta)$. Da somit die ganze Oberschranke fast sicher gegen die Unterschranke null konvergiert, muss auch der eingeschachtelte Term $|f_n(\theta_n^*) - f(\theta)|$ fast sicher gegen null konvergieren. Daraus folgt schließlich (3.39).

3.4 Plug-In-Schätzer bei unbekannten Risikoniveaus

Kommen wir zurück zu dem Fall, dass neben dem Bruchpunkt θ auch die Risikoniveaus a und b unbekannt sind. Hierfür schlagen Ferger & Klotsche (2009, S. 111-119)⁶² folgendes grobes Procedere vor:

- 1) Schätze zunächst die Risikoniveaus a und b .
- 2) Verwende anschließend den ML-Schätzer für θ aus Abschnitt 3.3, wobei die Schätzwerte aus 1) als Eingangsparameter verwendet werden.

⁶⁰Die Abschätzung nach unten ist trivial, die Abschätzung nach oben benutzt die Dreiecksungleichung. Vgl. dazu z. B. Bronstein et al. (2001, S. 30) oder Heuser (1994, S. 83).

⁶¹Vgl. z. B. van der Vaart (1998, S. 7, Theorem 2.3), Karr (1993, S. 148, Theorem 5.23) oder Serfling (1980, S. 24).

⁶²Diese Arbeit basiert auf Ferger (2009, S. 50-59). Für Quellenangaben wird deshalb der Einfachheit halber nur auf Ferger & Klotsche (2009) verwiesen, es sei denn, in Ferger (2009) befinden sich ausführlichere Informationen.

Es zeigt sich, dass sich der resultierende Zwei-Schritt-Plug-In-Schätzer für θ asymptotisch wie der ML-Schätzer verhält. Ferner lässt sich eine asymptotische Konfidenzregion für das unbekannte Tripel (θ, a, b) konstruieren.

3.4.1 Schätzung der Risikoniveaus

Unter der (vorübergehenden) Annahme, dass der Bruchpunkt θ bekannt sei, ergeben sich für die Risikoniveaus a und b die **Maximum-Likelihood-Schätzer**⁶³

$$a_n^* = a_n^*(\theta) = \frac{\sum_{i=1}^n Y_i \mathbb{1}_{\{X_i \leq \theta\}}}{\sum_{i=1}^n \mathbb{1}_{\{X_i \leq \theta\}}}, \quad X_{1:n} \leq \theta, \quad (3.42)$$

$$b_n^* = b_n^*(\theta) = \frac{\sum_{i=1}^n Y_i \mathbb{1}_{\{X_i > \theta\}}}{\sum_{i=1}^n \mathbb{1}_{\{X_i > \theta\}}}, \quad X_{n:n} > \theta. \quad (3.43)$$

Falls die Bedingungen $X_{1:n} \leq \theta$ bzw. $X_{n:n} > \theta$ nicht erfüllt sind, sind die entsprechenden Schätzwerte beliebig. Der Einfachheit halber werden sie dann auf null gesetzt.

Inhaltlich sind die resultierenden ML-Schätzwerte relative Ereignishäufigkeiten⁶⁴ für die durch den Bruchpunkt θ getrennten Teilstichproben. Erkennbar wird dies z. B. an der Formel für a_n^* , in der der Nenner $\sum_{i=1}^n \mathbb{1}_{\{X_i \leq \theta\}}$ die Anzahl der Merkmalsträger mit einer Merkmalsausprägung kleiner oder gleich θ angibt und der Zähler $\sum_{i=1}^n Y_i \mathbb{1}_{\{X_i \leq \theta\}}$ unter diesen die Anzahl der Erfolge ($Y_i = 1$) zählt.

Eine alternative, handlichere Darstellung der ML-Schätzer ist mit Hilfe der zufälligen empirischen Verteilungsfunktion F_n gemäß (2.1) und der zufälligen markierten empirischen Verteilungsfunktion H_n gemäß (3.12) möglich. Es gilt

$$a_n^* = \frac{H_n(\theta)}{F_n(\theta)}, \quad X_{1:n} \leq \theta, \quad (3.44)$$

$$b_n^* = \frac{\bar{Y} - H_n(\theta)}{1 - F_n(\theta)}, \quad X_{n:n} > \theta. \quad (3.45)$$

Formel (3.44) ergibt sich offensichtlich aus (3.42) durch Erweiterung mit dem Faktor $\frac{1}{n}$. Zur Herleitung von (3.45) aus (3.43) beachte $\mathbb{1}_{\{X_i > \theta\}} = 1 - \mathbb{1}_{\{X_i \leq \theta\}}$. Daher gilt

$$\begin{aligned} b_n^* &= \frac{\sum_{i=1}^n Y_i \mathbb{1}_{\{X_i > \theta\}}}{\sum_{i=1}^n \mathbb{1}_{\{X_i > \theta\}}} = \frac{\frac{1}{n} \sum_{i=1}^n Y_i (1 - \mathbb{1}_{\{X_i \leq \theta\}})}{\frac{1}{n} \sum_{i=1}^n (1 - \mathbb{1}_{\{X_i \leq \theta\}})} \\ &= \frac{\frac{1}{n} \sum_{i=1}^n Y_i - \frac{1}{n} \sum_{i=1}^n Y_i \mathbb{1}_{\{X_i \leq \theta\}}}{\frac{1}{n} \sum_{i=1}^n 1 - \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{\{X_i \leq \theta\}}} = \frac{\bar{Y} - H_n(\theta)}{1 - F_n(\theta)}. \end{aligned}$$

Anhand der Schreibweisen in (3.44) und (3.45) lässt sich außerdem gut zeigen, dass die Bedingungen $X_{1:n} \leq \theta$ bzw. $X_{n:n} > \theta$ f. s. für schließlich alle $n \in \mathbb{N}$ gelten und die ML-

⁶³Vgl. Ferger & Klotsche (2009, S. 111) oder – ausführlicher – Ferger (2009, S. 50-51).

⁶⁴In Anwendungen sind das Ausfallquoten, Erkrankungsquoten usw.

Schätzer somit wohldefiniert sind: Aus dem Satz von Gliwenko⁶⁵ folgt $F_n(\theta) \xrightarrow{\text{f. s.}} F(\theta)$. Wegen $F(\theta) \in]0, 1[$ existiert f. s. ein $n_0 \in \mathbb{N}$, so dass für alle $n \geq n_0$ die Beziehung $0 < F_n(\theta) < 1$ erfüllt ist und daher $X_{1:n} \leq \theta$ und $X_{n:n} > \theta$ gelten.

Man kann die Gleichungen (3.42) bis (3.45) auch als Funktionen von θ auffassen. Werden die Funktionen mit den Argumenten s bzw. t versehen, ergeben sich

$$a_n^*(s) = \frac{\sum_{i=1}^n Y_i \mathbb{1}_{\{X_i \leq s\}}}{\sum_{i=1}^n \mathbb{1}_{\{X_i \leq s\}}} = \frac{H_n(s)}{F_n(s)}, \quad X_{1:n} \leq s \quad (3.46)$$

und

$$b_n^*(t) = \frac{\sum_{i=1}^n Y_i \mathbb{1}_{\{X_i > t\}}}{\sum_{i=1}^n \mathbb{1}_{\{X_i > t\}}} = \frac{\bar{Y} - H_n(t)}{1 - F_n(t)}, \quad X_{n:n} > t. \quad (3.47)$$

Im Folgenden wird gezeigt, dass diese Funktionen fast sicher punktweise konvergieren. Damit lassen sich unter gewissen Bedingungen an s und t stark konsistente und asymptotisch normalverteilte Schätzer für die Risikoniveaus a und b angeben.

Satz 3.31 (Starke Konsistenz, asymptotische Normalität)

Gegeben sei das Modell (3.1) mit einstufiger Regressionsfunktion κ wie in (3.2), wobei $a, b \in]0, 1[$, $a \neq b$ und $\theta \in \mathbb{R}$ gelte. Zudem gelte $(X_i, Y_i) \stackrel{\text{i.i.d.}}{\sim} (X, Y)$, $i = 1, \dots, n$. Des Weiteren seien $s \in \mathbb{R}$ mit $F(s) > 0$ und $t \in \mathbb{R}$ mit $F(t) < 1$ gegeben. Dann gelten

$$a_n^*(s) \xrightarrow{\text{f. s.}} a(s) := \begin{cases} a, & \text{falls } s \leq \theta, \\ \frac{aF(\theta) + b(F(s) - F(\theta))}{F(s)}, & \text{falls } s > \theta, \end{cases}$$

und

$$b_n^*(t) \xrightarrow{\text{f. s.}} b(t) := \begin{cases} \frac{a(F(\theta) - F(t)) + b(1 - F(\theta))}{1 - F(t)}, & \text{falls } t \leq \theta, \\ b, & \text{falls } t > \theta. \end{cases}$$

Gilt zusätzlich $s \leq \theta \leq t$, dann folgen

$$(a_n^*(s), b_n^*(t)) \xrightarrow{\text{f. s.}} (a, b) \quad (3.48)$$

und

$$\sqrt{n}\{(a_n^*(s), b_n^*(t)) - (a, b)\} \xrightarrow{d} (U, V), \quad (3.49)$$

⁶⁵Der Satz von Gliwenko(-Cantelli), auch Hauptsatz der Mathematischen Statistik, bestätigt die f. s. gleichmäßige Konvergenz der empirischen Verteilungsfunktion F_n gegen die zugehörige theoretische Verteilungsfunktion F . Vgl. z. B. Karr (1993, S. 206), Müller (1991, S. 84-85) oder Rinne (2008, S. 425).

wobei $U \sim N(0, \sigma_U^2(s))$ und $V \sim N(0, \sigma_V^2(t))$ stochastisch unabhängig sind und die Varianzen

$$\sigma_U^2(s) = \frac{a(1-a)}{F(s)} \quad \text{und} \quad \sigma_V^2(t) = \frac{b(1-b)}{1-F(t)} \quad (3.50)$$

besitzen.

Beweis: Vgl. Ferger & Klotsche (2009, S. 112), für (3.48) und (3.49) insbesondere die Propositionen 5.1 bzw. 5.2. Beachte dabei $b(\theta) = b$. \square

Die Aussagen (3.48) und (3.49) zeigen insbesondere die starke Konsistenz und die asymptotische Normalität⁶⁶ der ML-Schätzer für die Risikoniveaus a und b unter der Annahme, dass der Bruchpunkt θ bekannt ist:

$$(a_n^*, b_n^*) \xrightarrow{\text{f. s.}} (a, b),$$

$$\sqrt{n}\{(a_n^*, b_n^*) - (a, b)\} \xrightarrow{\text{d}} (U, V).$$

Hinreichend für die besagten Eigenschaften ist aber bereits die Kenntnis eines Intervalls $[s, t]$, welches den unbekannten Parameter θ enthält. Je schmaler dieses Intervall ist, desto besser werden die Schätzer in dem Sinne, dass die Varianzen in (3.50) kleiner werden. Beide Varianzen werden minimal für $s = \theta = t$, d. h. wenn der Bruchpunkt bekannt ist.⁶⁷ Sind im Gegensatz dazu keine oder kaum Informationen über die Lage des Bruchpunkts verfügbar, muss das Intervall $[s, t]$ sehr groß gewählt werden. Die Varianzen können dadurch sehr groß werden. Unter der Annahme, dass $F(s) \neq 0$ für alle $s \in \mathbb{R}$ und $F(t) \neq 1$ für alle $t \in \mathbb{R}$,⁶⁸ gilt

$$s \rightarrow -\infty \Rightarrow F(s) \downarrow 0 \Rightarrow \sigma_U^2(s) \rightarrow +\infty,$$

$$t \rightarrow +\infty \Rightarrow F(t) \uparrow 1 \Rightarrow \sigma_V^2(t) \rightarrow +\infty.$$

Da meist ungenügend Informationen über die Lage des Bruchpunkts θ vorliegen, schlagen Ferger & Klotsche (2009, S. 112) Plug-In-Schätzer (Substitutionsschätzer) für die Risikoniveaus a und b vor: Man ersetze in (3.42) und (3.43) den Parameter θ jeweils durch einen Schätzer, der keine Kenntnis von a und b verlangt. Ferger & Klotsche wählen den zweiseitigen Dempfle-Stute-Schätzer $\bar{\theta}_n$ und erhalten damit die **Plug-In-Schätzer**

$$\bar{a}_n^* := a_n^*(\bar{\theta}_n) \quad \text{und} \quad \bar{b}_n^* := b_n^*(\bar{\theta}_n) \quad (3.51)$$

für die Risikoniveaus a und b .⁶⁹

⁶⁶Zum Begriff der asymptotischen Normalität einer Folge von Zufallsvektoren vgl. z. B. Serfling (1980, S. 21). Mit dem sogenannten Cramér-Wold-Device (vgl. z. B. Karr, 1993, S. 150) folgt daraus die asymptotische Normalität der einzelnen Komponenten des Vektors.

⁶⁷Vgl. auch im Folgenden Ferger & Klotsche (2009, S. 112).

⁶⁸Das ist z. B. der Fall, wenn die Verteilungsfunktion F auf ganz \mathbb{R} streng monoton wachsend ist.

⁶⁹Vgl. auch Ferger (2009, S. 53).

Bemerkung 3.32 (Version des Dempfle-Stute-Schätzers)

- (i) *An dieser Stelle sei erwähnt, dass die im Folgenden vorgestellten Aussagen nicht nur für die Plug-In-Schätzer gelten, die auf dem zweiseitigen Dempfle-Stute-Schätzer $\bar{\theta}_n$ beruhen. Wenn $a > b$ oder $a < b$ bekannt ist, dann gelten alle Sätze und Folgerungen des Abschnitts 3.4 ebenso für die Substitutionsschätzer, die auf den zugehörigen einseitigen Dempfle-Stute-Schätzern $\hat{\theta}_n$ bzw. $\hat{\theta}_n^-$ basieren.*
- (ii) *Wenn die Relation der Risikoniveaus a und b , d. h. entweder $a > b$ oder $a < b$, bekannt ist, dann sollte die entsprechende einseitige Variante des Dempfle-Stute-Schätzers genutzt werden. Diese stellt im nichtdegenerierten Fall sicher, dass die geschätzten Risikoniveaus ebenso die unterstellte Relation erfüllen:*

$$\begin{aligned} \text{Es sei } a > b. \text{ Dann gilt: } S_n(\hat{\theta}_n) > 0 &\Rightarrow a_n^*(\hat{\theta}_n) > b_n^*(\hat{\theta}_n). \\ \text{Es sei } a < b. \text{ Dann gilt: } S_n(\hat{\theta}_n^-) < 0 &\Rightarrow a_n^*(\hat{\theta}_n^-) < b_n^*(\hat{\theta}_n^-). \end{aligned}$$

Im degenerierten Fall, d. h. beispielsweise bei $S_n(\hat{\theta}_n) \leq 0$ unter der Annahme $a > b$,⁷⁰ und bei der zweiseitigen Variante des Dempfle-Stute-Schätzers überträgt sich die Relation der Risikoniveaus nicht automatisch auf die Relation der zugehörigen Schätzer.

Ursächlich für diese Zusammenhänge sind die folgenden Äquivalenzen. Für alle x mit $F_n(x) \in]0, 1[$ gilt gemäß (3.14) sowie (3.46) und (3.47):

$$\begin{aligned} S_n(x) = H_n(x) - \bar{Y}F_n(x) > 0 &\Leftrightarrow \bar{Y} - \bar{Y}F_n(x) > \bar{Y} - H_n(x) \\ \Updownarrow & \qquad \qquad \qquad \Updownarrow \\ a_n^*(x) = \frac{H_n(x)}{F_n(x)} > \bar{Y} & \qquad \qquad \qquad b_n^*(x) = \frac{\bar{Y} - H_n(x)}{1 - F_n(x)} < \bar{Y} \end{aligned}$$

und analog

$$S_n(x) < 0 \Leftrightarrow a_n^*(x) < \bar{Y} < b_n^*(x).$$

Im Folgenden konzentrieren wir uns auf die Plug-In-Schätzer aus (3.51), die auf der zweiseitigen Variante des Dempfle-Stute-Schätzers beruhen, weil sie den allgemeinsten Fall abdecken. Unter gewissen Annahmen zur Verteilungsfunktion F lassen sich für die Schätzer starke Konsistenz und asymptotische Normalität zeigen.

⁷⁰In diesem Fall ist $\hat{\theta}_n = X_{1:n}$, weil S_n nirgends positiv ist.

Satz 3.33 (Starke Konsistenz, asymptotische Normalität)

Gegeben sei das Modell (3.1) mit einstufiger Regressionsfunktion κ wie in (3.2), wobei $a, b \in]0, 1[$, $a \neq b$ und $\theta \in \mathbb{R}$ gelte. Zudem gelte $(X_i, Y_i) \stackrel{\text{i.i.d.}}{\sim} (X, Y)$, $i = 1, \dots, n$.

- (i) Ist die Verteilungsfunktion F für ein beliebiges $r > 0$ stetig und streng monoton wachsend in einer Umgebung $B_r(\theta)$, dann gilt

$$(\bar{a}_n^*, \bar{b}_n^*) \xrightarrow{\text{f. s.}} (a, b).$$

- (ii) Ist die Verteilungsfunktion F in einer (beliebig kleinen) Umgebung von θ stetig differenzierbar mit $F'(\theta) > 0$, dann gilt

$$\sqrt{n}\{(\bar{a}_n^*, \bar{b}_n^*) - (a, b)\} \xrightarrow{d} (U, V), \quad (3.52)$$

wobei $U \sim N(0, \sigma_U^2)$ und $V \sim N(0, \sigma_V^2)$ stochastisch unabhängig sind und die Varianzen

$$\sigma_U^2 := \sigma_U^2(\theta) = \frac{a(1-a)}{F(\theta)} \quad \text{und} \quad \sigma_V^2 := \sigma_V^2(\theta) = \frac{b(1-b)}{1-F(\theta)} \quad (3.53)$$

besitzen.

Beweis:

- (i) Vgl. Ferger & Klotsche (2009, S. 113, Proposition 5.3). Die Annahme der strengen Monotonie fehlt in Proposition 5.3. Sie ist aber nötig, damit die Aussage $\bar{\theta}_n \xrightarrow{\text{f. s.}} \theta$, welche im Beweis genutzt wird, gesichert ist. Vgl. dazu Satz 3.12. Mit dieser zusätzlichen Voraussetzung ist die Bedingung $0 < F(\theta) < 1$ aus der Proposition redundant. Vgl. dazu Bemerkung 3.13.
- (ii) Vgl. Ferger & Klotsche (2009, S. 113, Proposition 5.4). Die Bedingung $0 < F(\theta) < 1$ aus der Proposition ist redundant. Vgl. dazu Bemerkung 3.18(ii) in Verbindung mit Bemerkung 3.13.

□

Bemerkung 3.34 (Zu den Voraussetzungen von Satz 3.33(i))

Anstelle die strenge Monotonie von F auf einer Umgebung von θ vorauszusetzen, wäre es auch möglich gewesen, die Gültigkeit von Annahme 3.21 zu fordern. Diese beiden Bedingungen sind nämlich unter der zusätzlich getätigten Prämisse lokaler Stetigkeit von F äquivalent.

3.4.2 Schätzung des Bruchpunkts

Der ML-Schätzer θ_n^* aus Abschnitt 3.3 lässt sich wegen seiner Abhängigkeit von den Risikoniveaus a und b exakter als

$$\theta_n^* = \theta_n^*(a, b)$$

schreiben. Wenn a und b unbekannt sind, erscheint es auf Basis von Abschnitt 3.4.1 nun sinnvoll, die nichtparametrischen Plug-In-Schätzer \bar{a}_n^* und \bar{b}_n^* als Grundlage für den ML-Schätzer zu verwenden. Es resultiert der ebenfalls nichtparametrische **Zwei-Schritt-Plug-In-Schätzer**⁷¹

$$\bar{\theta}_n^* := \theta_n^*(\bar{a}_n^*, \bar{b}_n^*) = \theta_n^*(a_n^*(\bar{\theta}_n), b_n^*(\bar{\theta}_n)).$$

Es ist

$$\bar{\theta}_n^* = \operatorname{argmax}_{x \in \mathbb{R}} \bar{S}_n^*(x),$$

wobei

$$\bar{S}_n^*(x) := \bar{\alpha}_n^* H_n(x) + \bar{\beta}_n^* F_n(x) \quad (3.54)$$

mit

$$\bar{\alpha}_n^* := \ln \frac{\bar{a}_n^*(1 - \bar{b}_n^*)}{\bar{b}_n^*(1 - \bar{a}_n^*)} \quad \text{und} \quad \bar{\beta}_n^* := \ln \frac{(1 - \bar{a}_n^*)}{(1 - \bar{b}_n^*)}.$$

Der Schätzer $\bar{\theta}_n^*$ ist nicht definiert, wenn $\bar{\alpha}_n^*$ und/oder $\bar{\beta}_n^*$ nicht definiert sind. Das ist dann der Fall, wenn \bar{a}_n^* und/oder \bar{b}_n^* einen der Werte 0 oder 1 annehmen.

Beispiel 3.35 (Fortsetzung von Beispiel 3.8, 3.10)

Im obigen Beispiel⁷² ist $\hat{\theta}_n = 1$. Gemäß (3.22) ist $\bar{\theta}_n = \hat{\theta}_n = 1$. Mit (3.51) sowie (3.46) und (3.47) folgen die Schätzwerte

$$\bar{a}_n^* = a_n^*(\bar{\theta}_n) = \frac{H_n(1)}{F_n(1)} = \frac{1/2}{1/2} = 1 \quad \text{und} \quad \bar{b}_n^* = b_n^*(\bar{\theta}_n) = \frac{\bar{y} - H_n(1)}{1 - F_n(1)} = \frac{1/2 - 1/2}{1 - 1/2} = 0.$$

Folglich existieren die Schätzwerte $\bar{\alpha}_n^*$ und $\bar{\beta}_n^*$ nicht (Division durch null bzw. Logarithmus von null) und der Schätzwert $\bar{\theta}_n^*$ ist nicht definiert.

Notwendig für die Existenz des Plug-In-Schätzwertes ist es, dass sowohl links als auch rechts von $\bar{\theta}_n$ mindestens zwei Beobachtungen des Merkmals X liegen und dass sich unter den zugehörigen Y -Beobachtungen jeweils mindestens ein Erfolg und ein Misserfolg befinden. Unter den Voraussetzungen von Satz 3.33(i) existiert der Substitutionsschätzer $\bar{\theta}_n^*$ f. s. für schließlich alle $n \in \mathbb{N}$, da die unbekannten Parameter a und b verschieden von null und eins sind.

⁷¹Vgl. Ferger & Klotsche (2009, S. 114-115).

⁷²Siehe S. 47 und S. 50.

Der Zwei-Schritt-Plug-In-Schätzer $\bar{\theta}_n^*$ zeigt ebenso wie der ML-Schätzer wünschenswerte asymptotische Eigenschaften. So werden in den folgenden Sätzen Aussagen über die starke Konsistenz und die n -Konsistenz von $\bar{\theta}_n^*$ gegen θ sowie über die fast schwache Konvergenz der Folge $n(\bar{\theta}_n^* - \theta)$ bewiesen.

Satz 3.36 (Starke Konsistenz)

Gegeben sei das Modell (3.1) mit einstufiger Regressionsfunktion κ wie in (3.2), wobei $a, b \in]0, 1[$, $a \neq b$ und $\theta \in \mathbb{R}$ gelte. Zudem gelte $(X_i, Y_i) \stackrel{\text{i.i.d.}}{\sim} (X, Y)$, $i = 1, \dots, n$. Falls Annahme 3.21 gilt und F zusätzlich stetig in einer Umgebung von θ ist, dann ist der Substitutionsschätzer $\bar{\theta}_n^*$ stark konsistent für θ , d. h.

$$\bar{\theta}_n^* \xrightarrow{\text{f. s.}} \theta \text{ für alle } \theta \in \mathbb{R}.$$

Beweis: Vgl. Ferger & Klotsche (2009, S. 115, Proposition 5.5). □

Bemerkung 3.37 (Zu den Voraussetzungen von Satz 3.36)

- (i) Die Annahme des Satzes bezüglich der Verteilungsfunktion F ist unter Beachtung von Bemerkung 3.34 identisch mit der Voraussetzung von Satz 3.33(i). Daher sind die Plug-In-Schätzer \bar{a}_n^* und \bar{b}_n^* stark konsistent. Somit existiert der Substitutionsschätzer $\bar{\theta}_n^*$ f. s. für schließlich alle $n \in \mathbb{N}$ und asymptotische Aussagen sind sinnvoll.
- (ii) Die Annahme des Satzes bezüglich der Verteilungsfunktion F impliziert offensichtlich die Annahme des Satzes 3.22 zur starken Konsistenz des ML-Schätzers bei bekannten Risikoniveaus. In Anbetracht von Bemerkung 3.34 ist sie ferner identisch mit den Voraussetzungen des Satzes 3.12 zur starken Konsistenz der Dempfle-Stute-Schätzer.

Satz 3.38 (optimale Konvergenzgeschwindigkeit, n -Konsistenz)

Gegeben sei das Modell (3.1) mit einstufiger Regressionsfunktion κ wie in (3.2), wobei $a, b \in]0, 1[$, $a \neq b$ und $\theta \in \mathbb{R}$ gelte. Zudem gelte $(X_i, Y_i) \stackrel{\text{i.i.d.}}{\sim} (X, Y)$, $i = 1, \dots, n$. Falls es ein $r > 0$ und positive Konstanten $\underline{L} = \underline{L}(r)$ und $\bar{L} = \bar{L}(r)$ derart gibt, dass

$$\underline{L}|s - t| \leq |F(s) - F(t)| \leq \bar{L}|s - t| \text{ für alle } s, t \in B_r(\theta) \quad (3.55)$$

gilt, dann gilt

$$n(\bar{\theta}_n^* - \theta) = \mathcal{O}_P(1).$$

Beweis: Die Behauptung folgt aus Proposition 5.6 in Ferger & Klotsche (2009, S. 115) mittels des Grenzübergangs $x \rightarrow \infty$.⁷³ Die weiteren Annahmen aus Proposition 5.6,

⁷³Vgl. auch Ferger (2009, S. 57, Satz 5.17).

d. h. die Stetigkeit und die strenge Monotonie von F auf einer Umgebung von θ , sind redundant. Sie folgen aus der Bedingung (3.55). Vgl. dazu Bemerkung 3.24(i) und (ii) zur inhaltlich identischen Annahme (3.30). \square

Bemerkung 3.39 (Zur Voraussetzung von Satz 3.38)

Die Annahme des Satzes bezüglich der Verteilungsfunktion F ist identisch mit der Voraussetzung des Satzes 3.23 zur n -Konsistenz des ML-Schätzers bei bekannten Risikoniveaus. Die Voraussetzung des Satzes 3.16 zur n -Konsistenz der Dempfle-Stute-Schätzer ist stärker, d. h. sie impliziert (3.55), wie folgendes Lemma zeigt.

Lemma 3.40 Gegeben sei ein $r > 0$. Ist die Verteilungsfunktion F auf dem Intervall $B_r(\theta) =]\theta - r, \theta + r[$ stetig differenzierbar mit $F'(\theta) > 0$, dann existieren ein $r^* > 0$ und positive Konstanten $\underline{L} = \underline{L}(r^*)$ und $\bar{L} = \bar{L}(r^*)$ derart, dass für alle $s, t \in B_{r^*}(\theta) =]\theta - r^*, \theta + r^*[$ die Aussage

$$\underline{L}|s - t| \leq |F(s) - F(t)| \leq \bar{L}|s - t|$$

gilt.

Beweis: Wegen der Voraussetzung der stetigen Differenzierbarkeit von F sind sowohl F als auch die Ableitung F' auf $B_r(\theta)$ stetig. Weil F' insbesondere in θ stetig ist und $0 < \frac{1}{2}F'(\theta) < F'(\theta) < F'(\theta) + 1$ gilt, existiert ein $0 < r^* \leq r$, so dass für alle $x \in B_{r^*}(\theta)$ gilt:⁷⁴

$$0 < \frac{1}{2}F'(\theta) < F'(x) < F'(\theta) + 1. \quad (3.56)$$

Gegeben seien nun ohne Beschränkung der Allgemeinheit $t < s \in B_{r^*}(\theta)$.⁷⁵ Auf dem kompakten Intervall $[t, s] \subset B_{r^*}(\theta) \subseteq B_r(\theta)$ ist F stetig und differenzierbar. Gemäß dem Mittelwertsatz der Differentialrechnung⁷⁶ gibt es daher ein $\zeta \in]t, s[$ mit

$$F'(\zeta) = \frac{F(s) - F(t)}{s - t}.$$

Weil F als Verteilungsfunktion monoton wachsend ist und weil $\zeta \in B_{r^*}(\theta)$ ist, folgt daraus mit (3.56) die Ungleichungskette

$$0 < \frac{1}{2}F'(\theta) < F'(\zeta) = \frac{F(s) - F(t)}{s - t} = \frac{|F(s) - F(t)|}{|s - t|}.$$

Dies zeigt den ersten Teil der Behauptung mit $\underline{L} := \frac{1}{2}F'(\theta) > 0$.

⁷⁴Vgl. Heuser (1994, S. 214, Satz 34.2). Wähle dabei $X = B_r(\theta)$.

⁷⁵Für $s = t$ gilt die Behauptung des Lemmas trivialerweise. Im Falle $s < t$ sind die Rollen zu tauschen.

⁷⁶Vgl. z. B. Heuser (1994, S. 279).

Der zweite Teil der Behauptung folgt aus Satz 49.4 in Heuser (1994, S. 280): Weil F gemäß (3.56) auf dem Intervall $B_{r^*}(\theta)$ eine beschränkte Ableitung besitzt, ist F auf diesem Intervall sogar dehnungsbeschränkt (Lipschitz-stetig) mit der Dehnungsschranke $\bar{L} := \sup_{x \in B_{r^*}(\theta)} |F'(x)|$, d. h.⁷⁷

$$|F(s) - F(t)| \leq \bar{L}|s - t| \quad \text{für alle } s, t \in B_{r^*}(\theta).$$

Ferner gilt $\bar{L} = \sup_{x \in B_{r^*}(\theta)} |F'(x)| \geq F'(\theta) > 0$. Damit ist das Lemma vollständig bewiesen. \square

Das folgende Resultat zur fast schwachen Konvergenz der Folge $n(\bar{\theta}_n^* - \theta)$ zeigt abschließend – wie am Anfang des Abschnitts 3.4 angekündigt –, dass sich der Plug-In-Schätzer $\bar{\theta}_n^*$ mit geschätzten Parametern a und b und der ML-Schätzer θ_n^* bei bekannten Parametern a und b asymptotisch gleich verhalten. Unter den gleichen Voraussetzungen haben nämlich beide Folgen im Sinne der fast schwachen Konvergenz dieselben Grenzvarenablen.

Satz 3.41 (Fast schwache Konvergenz)

Gegeben sei das Modell (3.1) mit einstufiger Regressionsfunktion κ wie in (3.2), wobei $a, b \in]0, 1[$, $a \neq b$ und $\theta \in \mathbb{R}$ gelte. Zudem gelte $(X_i, Y_i) \stackrel{\text{i.i.d.}}{\sim} (X, Y)$, $i = 1, \dots, n$. Falls die Verteilungsfunktion F die Bedingung (3.55) erfüllt und an der Stelle θ eine rechtsseitige Ableitung $F'(\theta+)$ sowie eine linksseitige Ableitung $F'(\theta-)$ besitzt, dann gelten

$$\limsup_{n \rightarrow \infty} P(n(\bar{\theta}_n^* - \theta) \leq x) \leq P(\tau \leq x)$$

und

$$\liminf_{n \rightarrow \infty} P(n(\bar{\theta}_n^* - \theta) < x) \geq P(\sigma < x)$$

für alle $x \in \mathbb{R}$, d. h.

$$n(\bar{\theta}_n^* - \theta) \rightsquigarrow_{a.w.} (\tau, \sigma). \quad (3.57)$$

Dabei bezeichnen τ und σ wiederum die kleinste bzw. größte Maximalstelle des zusammengesetzten Poisson-Prozesses L aus (3.34).

Beweis: Vgl. Ferger & Klotsche (2009, S. 117, Theorem 5.8). \square

Als Folge der fast schwachen Konvergenz gemäß (3.57) lässt sich ein asymptotisches Konfidenzintervall für θ auf Basis des Plug-In-Schätzers angeben.

⁷⁷Vgl. Heuser (1994, S. 136 u. 212). Siehe auch Bemerkung 3.24(i).

Folgerung 3.42 (Asymptotisches Konfidenzintervall für θ)

Gegeben seien die Voraussetzungen und Bezeichnungen von Satz 3.41. Für das Paar (c_1, d_1) gelte $P(\sigma < c_1) - P(\tau \leq -d_1) \geq 1 - \gamma$, wobei $0 < \gamma < 1$. Dann ist das Intervall

$$I_n^*(\gamma) = \left] \bar{\theta}_n^* - \frac{c_1}{n}, \bar{\theta}_n^* + \frac{d_1}{n} \right[$$

ein asymptotisches Konfidenzintervall für θ zum Konfidenzniveau $1 - \gamma$, d. h.

$$\liminf_{n \rightarrow \infty} P(\theta \in I_n^*(\gamma)) \geq 1 - \gamma.$$

Beweis: Vgl. Ferger & Klotsche (2009, S. 117). Dort wird das Konfidenzniveau mit $1 - \alpha$ bezeichnet. Diese Notation ist nicht zu verwechseln mit der Größe α , welche aus den Risikoniveaus a und b gebildet wird. \square

Zur Bestimmung der Zahlen c_1 und d_1 empfehlen Ferger & Klotsche (2009, S. 117-118) eine Monte-Carlo-Approximation auf Basis der expliziten Darstellungen der Zufallsvariablen τ und σ gemäß Bemerkung 3.29. Dabei werden die Parameter a , b , $\lambda_1 = F'(\theta+)$ und $\lambda_2 = F'(\theta-)$ benötigt. Diese sind allesamt unbekannt und müssen daher durch Schätzwerte ersetzt werden. Ferger & Klotsche schlagen den Einsatz des Quadrupels $(\bar{a}_n^*, \bar{b}_n^*, \bar{\lambda}_n, \bar{\lambda}_n)$ vor. Dabei sind \bar{a}_n^* und \bar{b}_n^* die Plug-In-Schätzer für a bzw. b gemäß (3.51). Der Intensitätsschätzer $\bar{\lambda}_n := f_n(\bar{\theta}_n)$ basiert auf dem zweiseitigen Dempfle-Stute-Schätzer. Wie in Abschnitt 3.3 werden dabei die folgenden vereinfachenden Annahmen getroffen:

- Die Verteilungsfunktion F besitze die stetige Dichte f , so dass $\lambda_1 = \lambda_2 = f(\theta)$ gilt.
- Der Dichteschätzer f_n konvergiert auf einer Umgebung $B_r(\theta)$, $r > 0$ fast sicher gleichmäßig gegen f .

Diese Annahmen führen analog zu Bemerkung 3.30 zur starken Konsistenz des Intensitätsschätzers, d. h.

$$\bar{\lambda}_n = f_n(\bar{\theta}_n) \xrightarrow{\text{f. s.}} f(\theta).$$

Unerwähnt bleibt, warum der zweiseitige Dempfle-Stute-Schätzer als Grundlage für den Intensitätsschätzer gewählt wird. Ebenso möglich und unter den genannten Voraussetzungen an f bzw. f_n ebenso sinnvoll wäre der Intensitätsschätzer auf Basis des Zwei-Schritt-Plug-In-Schätzers, d. h.

$$\bar{\lambda}_n^* := f_n(\bar{\theta}_n^*) \xrightarrow{\text{f. s.}} f(\theta).$$

3.4.3 Simultane Schätzung

Mit (3.52) und (3.57) sind Verteilungsaussagen zu den drei nichtparametrischen Schätzern \bar{a}_n^* , \bar{b}_n^* und $\bar{\theta}_n^*$ verfügbar. Das Theorem 6.1 in Ferger & Klotsche (2009, S. 119) liefert eine simultane Erweiterung dieser Ergebnisse. Dabei werden

$$\bar{\sigma}_U^2 = \frac{\bar{a}_n^*(1 - \bar{a}_n^*)}{F_n(\bar{\theta}_n^*)} \quad \text{und} \quad \bar{\sigma}_V^2 = \frac{\bar{b}_n^*(1 - \bar{b}_n^*)}{1 - F_n(\bar{\theta}_n^*)} \quad (3.58)$$

als Schätzer für die Varianzen aus (3.53) verwendet.

Satz 3.43 (Asymptotisches Verhalten der gemeinsamen Verteilung)

Gegeben sei das Modell (3.1) mit einstufiger Regressionsfunktion κ wie in (3.2), wobei $a, b \in]0, 1[$, $a \neq b$ und $\theta \in \mathbb{R}$ gelte. Zudem gelte $(X_i, Y_i) \stackrel{\text{i.i.d.}}{\sim} (X, Y)$, $i = 1, \dots, n$. Des Weiteren sei die Verteilungsfunktion F in einer (beliebig kleinen) Umgebung von θ stetig differenzierbar mit $F'(\theta) > 0$. Sind zusätzlich $N_1 \sim N(0, 1)$ und $N_2 \sim N(0, 1)$ stochastisch unabhängig, dann gelten

$$\begin{aligned} \liminf_{n \rightarrow \infty} P \left(n(\bar{\theta}_n^* - \theta) < z, \frac{\sqrt{n}(\bar{a}_n^* - a)}{\bar{\sigma}_U} \in I_1, \frac{\sqrt{n}(\bar{b}_n^* - b)}{\bar{\sigma}_V} \in I_2 \right) \\ \geq P(\sigma < z)P(N_1 \in I_1)P(N_2 \in I_2) \end{aligned}$$

und

$$\begin{aligned} \limsup_{n \rightarrow \infty} P \left(n(\bar{\theta}_n^* - \theta) \leq z, \frac{\sqrt{n}(\bar{a}_n^* - a)}{\bar{\sigma}_U} \in I_1, \frac{\sqrt{n}(\bar{b}_n^* - b)}{\bar{\sigma}_V} \in I_2 \right) \\ \leq P(\tau \leq z)P(N_1 \in I_1)P(N_2 \in I_2) \end{aligned}$$

für alle $z \in \mathbb{R}$ und alle Intervalle $I_1, I_2 \subseteq \mathbb{R}$. Dabei bezeichnen τ und σ wiederum die kleinste bzw. größte Maximalstelle des zusammengesetzten Poisson-Prozesses L aus (3.34).

Beweis: Vgl. Ferger & Klotsche (2009, S. 119, Theorem 6.1) sowie Ferger (2009, Kapitel 6, S. 65-81), insbesondere Theorem 6.4 und die Bezeichnungen aus Theorem 6.3. Die Bedingung $0 < F(\theta) < 1$ aus Theorem 6.1 in Ferger & Klotsche (2009, S. 119) ist redundant. Vgl. dazu Bemerkung 3.18(ii) in Verbindung mit Bemerkung 3.13. \square

Die Behauptung des Satzes gilt laut Ferger & Klotsche auch dann, wenn die Schätzer $\bar{\sigma}_U$ und $\bar{\sigma}_V$ durch die wahren Parameter σ_U und σ_V ersetzt werden. Sie enthält die Aussagen (3.52) für $z \rightarrow \infty$ und (3.57) für $I_1 = I_2 = \mathbb{R}$ als Spezialfälle.⁷⁸ Als Folge von Satz 3.43 lässt sich eine asymptotische Konfidenzregion für das Tripel (θ, a, b) konstruieren.

⁷⁸Vgl. dazu auch Lemma 3.40.

Folgerung 3.44 (Asymptotische Konfidenzregion für (θ, a, b))

Gegeben seien die Voraussetzungen und Bezeichnungen von Satz 3.43. Für die reellen Zahlen c_1, c_2, c_3, d_1, d_2 und d_3 gelte⁷⁹

$$(P(\sigma < c_1) - P(\tau \leq -d_1))(\Phi(c_2) - \Phi(-d_2))(\Phi(c_3) - \Phi(-d_3)) \geq 1 - \gamma,$$

wobei $0 < \gamma < 1$. Dann ist der zufällige Quader (cuboid)

$$C_n(\gamma) = \left[\bar{\theta}_n^* - \frac{c_1}{n}, \bar{\theta}_n^* + \frac{d_1}{n} \right] \times \left[\bar{a}_n^* - \frac{\bar{\sigma}_U}{\sqrt{n}} c_2, \bar{a}_n^* + \frac{\bar{\sigma}_U}{\sqrt{n}} d_2 \right] \times \left[\bar{b}_n^* - \frac{\bar{\sigma}_V}{\sqrt{n}} c_3, \bar{b}_n^* + \frac{\bar{\sigma}_V}{\sqrt{n}} d_3 \right] \quad (3.59)$$

eine asymptotische Konfidenzregion für (θ, a, b) zum Konfidenzniveau $1 - \gamma$, d. h.

$$\liminf_{n \rightarrow \infty} P((\theta, a, b) \in C_n(\gamma)) \geq 1 - \gamma.$$

Beweis: Vgl. Ferger & Klotsche (2009, S. 119) oder, für eine ausführlichere Darstellung, Ferger (2009, S. 80-81). In beiden Quellen wird das Konfidenzniveau mit $1 - \alpha$ bezeichnet. Diese Notation ist nicht zu verwechseln mit der Größe α , welche aus den Risikoniveaus a und b gebildet wird. \square

3.5 Vergleich der Schätzer

Zum Vergleich der Schätzer verweisen wir abermals auf Ferger & Klotsche (2009, Kap. 7, S. 120-121). In ihrer Simulationsstudie vergleichen sie den einseitigen und zweiseitigen Dempfle-Stute-Schätzer mit dem ML-Schätzer und dem Zwei-Schritt-Plug-In-Schätzer. Dabei verwenden sie die Annahme $a < b$, verschiedene Stichprobenumfänge n und unterschiedliche Parameterkonstellationen (θ, a, b) . Die Regressorvariable X ist standardnormalverteilt, d. h. $F = \Phi$. Die Verteilungsfunktion Φ ist streng monoton wachsend und stetig differenzierbar. Ihre erste Ableitung $\Phi' = \varphi$ ist auf ganz \mathbb{R} positiv und beschränkt. Somit erfüllt Φ die Voraussetzungen aller einschlägigen Sätze und Folgerungen aus Kapitel 3.⁸⁰ Als Grundlage für den Vergleich der Schätzer dienen die empirische Verzerrung (*bias*) und der geschätzte RMSE (*root mean squared error*), d. h. die Wurzel aus der mittleren quadratischen Abweichung des Schätzwerts vom wahren Parameter θ . Je kleiner die Beträge dieser Größen werden, desto besser ist die Schätzung.

Bezüglich des Stichprobenumfangs und der Sprunghöhe der Regressionsfunktion, d. h. des Abstands $|a - b|$, zeigen die Schätzer das erwartete Verhalten:

⁷⁹Dabei bezeichnet Φ die Verteilungsfunktion der Standardnormalverteilung.

⁸⁰Beachte dabei auch Lemma 3.40.

- Bei wachsendem Stichprobenumfang wird die Schätzung besser: Verzerrung und RMSE werden typischerweise betragsmäßig kleiner. Dies spricht dafür, dass die Schätzer asymptotisch erwartungstreu (unverzerrt, *unbiased*) bzw. konsistent im quadratischen Mittel (MSE-konsistent) sind.⁸¹ Ferner bestätigt sich bei den großen Stichprobenumfängen, dass sich sowohl die Dempfle-Stute-Schätzer untereinander als auch ML- und Plug-In-Schätzer asymptotisch gleich verhalten.
- Bei einem großen Abstand der Risikoniveaus, hier gleich 0.6, ist der RMSE kleiner als bei einer geringeren Differenz, hier 0.3.⁸²

Bemerkung 3.45 (Verhalten bei extremen Risikoniveaus)

In Bezug auf die in Kapitel 4 folgende Anwendung wäre es interessant gewesen, auch extremere Risikoniveaus zu betrachten. Gemeint sind damit

- *Risikoniveaus, die sich nur geringfügig unterscheiden, z. B. $|a - b| = 0.01$, und*
- *Risikoniveaus, die sich nahe bei null oder nahe bei eins befinden, z. B. $a = 0.005$.*

Es muss davon ausgegangen werden, dass der Stichprobenumfang sehr groß sein muss, um

- *im ersten Fall eine hinreichende Güte der Schätzung zu gewährleisten und*
- *im zweiten Fall die Existenz des Plug-In-Schätzers für θ weitestgehend zu sichern. Dazu sind nämlich beiderseits des zugehörigen Dempfle-Stute-Schätzwerts sowohl Erfolge als auch Nichterfolge vonnöten.*

Nicht genau vorauszusehen ist das Verhalten der Schätzer, wenn die Lage des Bruchpunkts verändert wird. Ist der Bruchpunkt der Median der Standardnormalverteilung, d. h. $\theta = 0$, dann liegen die Fehler bei allen untersuchten Schätzern in derselben Größenordnung. Wird der wahre Parameter θ jedoch in Richtung eines oberen Quantils der Verteilung von X verschoben, dann zeigt sich, dass der Plug-In-Schätzer dem einseitigen wie zweiseitigen Dempfle-Stute-Schätzer überlegen ist. Der ML-Schätzer θ_n^* liefert insgesamt noch etwas bessere Schätzwerte als der Plug-In-Schätzer, läuft aber außer Konkurrenz, da er die Kenntnis der Risikoniveaus a und b voraussetzt.

Auffällig ist, dass die Verzerrung bei den drei nichtparametrischen Schätzern negativ ist, wenn der Bruchpunkt θ ein oberes Quantil ist. Ob sich dieses Verhalten umkehrt, wenn der Bruchpunkt negativ gewählt wird, oder ob die negative Verzerrung

⁸¹Vgl. zu diesen Begriffen z. B. Serfling (1980, S. 48) oder Rinne (2008, S. 445, 449).

⁸²Bei den Parameterkonstellationen $(\theta, a, b) = (1.05, 0.2, 0.8)$ und $(\theta, a, b) = (1.05, 0.25, 0.55)$ sind die ermittelten Werte für den einseitigen Dempfle-Stute-Schätzer vollkommen identisch. Dies könnte auf einen Copy&Paste-Fehler hindeuten. Aufgrund der obigen Bemerkungen zur asymptotischen Gleichheit der Dempfle-Stute-Schätzer ist in diesem Fall vermutlich der Block unter der Annahme $(\theta, a, b) = (1.05, 0.2, 0.8)$ falsch.

eher an der Definition des argmax-Funktionalen liegt, bleibt offen und untersuchenswert. Ebenfalls untersuchenswert ist die Fragestellung, inwiefern sich die Schätzwerte verändern, wenn der Plug-In-Schätzer iterativ angewendet wird. Dabei würden z. B. im ersten Iterationsschritt die Risikoniveaus auf Basis des Zwei-Schritt-Plug-In-Schätzers $\bar{\theta}_n^*$ geschätzt und anschließend auf der Grundlage dieser Schätzwerte der Drei-Schritt-Plug-In-Schätzwert für θ berechnet.

Diese Fragen sollen im Folgenden jedoch keine Rolle spielen. Vielmehr soll es darum gehen, die Schätzer im Kontext des Kreditrisikos anzuwenden und zu prüfen, inwiefern sie mit bereits vorgeschlagenen Methoden zur Ratingklassenbildung konkurrieren können. Dabei muss davon ausgegangen werden, dass der Abstand der Risikoniveaus a und b erheblich kleiner ist als in den in der Simulation untersuchten Parameterkonstellationen.

Kapitel 4

Anwendung

In diesem Kapitel werden die im Kapitel 3 vorgestellten Methoden der Bruchpunktschätzung an einem Datensatz einer deutschen Auskunftei erprobt. Ziel ist es, die Score-Skala in zwei Ratingklassen zu teilen, d. h. die Trennstelle zwischen den Klassen zu schätzen. Die Schätzung der Ratingklassengrenze erfolgt dabei in Abschnitt 4.2 zunächst quartalsweise aus Querschnittsdaten. Die im Zeitablauf ermittelten Schätzwerte werden mit den Ergebnissen ausgewählter Methoden zur Ratingklasseneinteilung aus Abschnitt 2.2 verglichen. Konfidenzintervalle für die Trennstelle zwischen den Ratingklassen sowie simultane Konfidenzregionen für die Trennstelle und die ratingklassenspezifischen Ausfallwahrscheinlichkeiten werden ebenfalls ermittelt. Anschließend wird in Abschnitt 4.3 der Bruchpunkt aus dem kompletten Datensatz, d. h. dem Pool aller Quartale, geschätzt. Doch bevor wir mit der Schätzung beginnen, wird in Abschnitt 4.1 der Auskunftei-Datensatz vorgestellt. Dabei wird auch auf die kritischen Annahmen des Modells eingegangen. Wegen der Betrachtung über mehrere Zeitpunkte erhalten die relevanten Größen im Folgenden einen Zeitindex t .

4.1 Datensatz und Modellannahmen

4.1.1 Einführung

Der verwendete Datensatz basiert auf einer anonymisierten Stichprobe aus dem Datenbestand einer deutschen Auskunftei. Die Stichprobe umfasst etwa 2.5 % der im ersten Quartal des Jahres 2005 bei der Auskunftei gemeldeten, voll geschäftsfähigen Bürger Deutschlands. Die in der Stichprobe enthaltenen Einzelpersonen wurden, soweit möglich, quartalsweise bis in das Quartal I-2000 zurückverfolgt (Rückwärtsergänzung) und, soweit möglich, bis zum Quartal III-2010 fortlaufend weiter betrachtet (Vorwärtsergänzung). Für die resultierenden 43 aufeinanderfolgenden Quartale enthält der Datensatz Informationen über das Ausfallverhalten und den bankenspezifischen Bonitätsscore der betreffenden Individuen. Dieser Score wurde auf Grundlage der in Abschnitt 2.1.1 skiz-

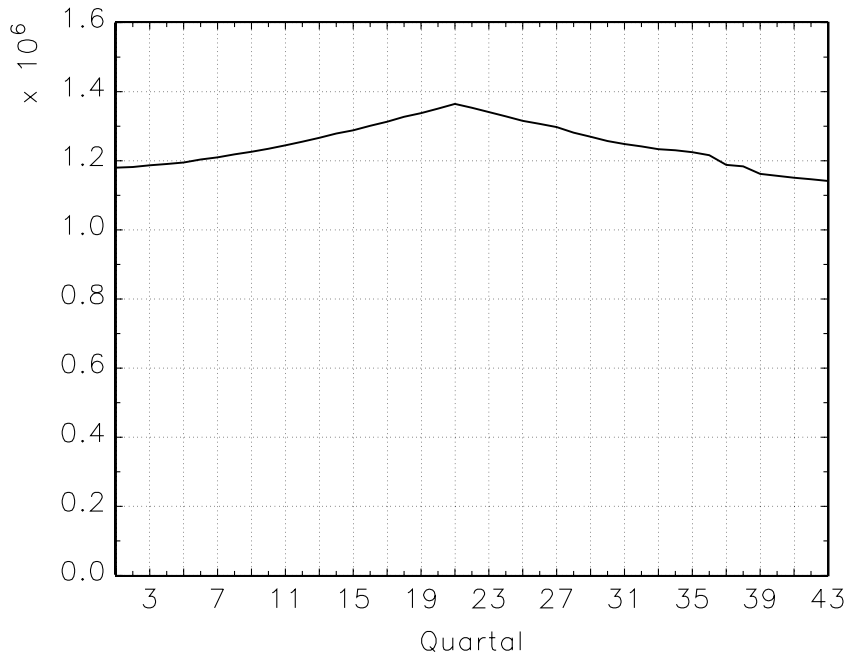


Abbildung 4.1: Anzahl der Beobachtungen im Zeitablauf

zierten logistischen Regression ermittelt. Er ist ganzzahlig und kann Werte zwischen 1 und 1000 annehmen. Zwischen Score und Bonität wird ein monotoner Zusammenhang unterstellt: Ein kleiner Score steht für eine schlechte Bonität und damit ein höheres Ausfallrisiko, umgekehrt repräsentiert ein großer Score eine gute Bonität und damit ein niedrigeres Ausfallrisiko. Speziell beinhaltet der Datensatz für jedes der 43 Quartale und für jeden der 1000 möglichen Scores

- die Anzahl der Individuen, die den entsprechenden Score am Anfang des betreffenden Quartals besaßen, und
- unter diesen die Anzahl der Ausfälle bis zum Ende des Quartals.

Ein Ausfall tritt beispielsweise dann ein, wenn ein Verbraucherinsolvenzverfahren eröffnet wird, ein Haftbefehl vorliegt oder eine eidesstattliche Versicherung abgegeben wird. Der Ausfallzustand ist nicht absorbierend, d. h. ein ausgefallenes Individuum kann wieder aufleben.

Abbildung 4.1 zeigt die Anzahl n_t , $t = 1, \dots, 43$ der Beobachtungen in der Stichprobe im Zeitablauf. Quartal 1 enthält die ältesten Daten aus dem Quartal I-2000. Von dort beginnend steigt der Stichprobenumfang, beispielsweise durch erstmalige Meldung bei der Auskunft, bis zum Stichtag in Quartal 21. Dieses besitzt mit $n_{21} = 1\,364\,419$ die größte Beobachtungszahl. Anschließend geht der Umfang der Stichprobe z. B. durch Tod oder dauerhaften Verlust der Geschäftsfähigkeit bis zum Quartal 43, welches die neuesten Daten (III-2010) enthält, immer weiter zurück. Wir betrachten somit eine Kohorte, die einer natürlichen Alterung unterliegt. Sollen im Folgenden gewisse Aspek-

te für einzelne Quartale vertieft dargestellt werden, fällt die Wahl stets auf die drei Quartale 1, 21 und 43. Sie markieren den Beginn, den Stichtag und das Ende des Untersuchungszeitraums.

Aus den Informationen des Datensatzes können Paare (x_{it}, y_{it}) , $i = 1, \dots, n_t$, $t = 1, \dots, 43$ von Score x_{it} und Ausfallindikator y_{it} gebildet werden. Der Score liegt wie bereits beschrieben in der Menge der natürlichen Zahlen von 1 bis 1000. Der Ausfallindikator nimmt den Wert 1 an, wenn das Ausfallereignis eingetreten ist, und sonst den Wert 0. Die Indizes i und t markieren das i -te Individuum im Quartal t . Dieses muss unter anderem wegen der im Zeitverlauf unterschiedlichen Stichprobenumfänge aber nicht mit dem i -ten Individuum in einem anderen Quartal übereinstimmen. Die Bonitäts- und Ausfallentwicklung einer speziellen Person kann also nicht über die Zeit verfolgt werden.

Die aus dem Auskunft-Datensatz generierten Paare (x_{it}, y_{it}) , $i = 1, \dots, n_t$, $t = 1, \dots, 43$ werden als Realisationen von Stichprobenvektoren (X_{it}, Y_{it}) , $i = 1, \dots, n_t$, $t = 1, \dots, 43$ aufgefasst. Es wird unterstellt, dass diese Zufallsvektoren stochastisch unabhängig und identisch verteilt sind. Für die quartalsbezogene Querschnittsbetrachtung heißt das also

$$(X_{it}, Y_{it}) \stackrel{\text{i.i.d.}}{\sim} (X_t, Y_t) \text{ für jedes feste } t = 1, \dots, 43.$$

Bezogen auf den gesamten Zeitverlauf lautet die Annahme

$$(X_{it}, Y_{it}) \stackrel{\text{i.i.d.}}{\sim} (X, Y).$$

Die i.i.d.-Annahme ist die Standardannahme der klassischen Statistik, jedoch im Kontext des Kreditrisikos unüblich. Hier ist gerade die Untersuchung und Modellierung der Abhängigkeitsstrukturen der Ausfall- oder Bonitätsvariablen von Interesse.¹ Ein gebräuchliches Element der Modellierungsideen ist die bedingte Unabhängigkeit der Variablen, d. h. bedingt auf die aktuell herrschende konjunkturelle Situation sind die Ausfall- oder Bonitätsvariablen stochastisch unabhängig.² Auf diese Sichtweise können wir uns in der quartalsbezogenen Betrachtung auch zurückziehen. Die resultierenden Schätzwerte sind dann entsprechend zu interpretieren. In der quartalsübergreifenden Untersuchung verwenden wir die i.i.d.-Annahme zu Referenzzwecken. Dies ist auch im Kreditrisiko nicht unüblich, da Kreditrisikomodelle durch die i.i.d.-Annahme einfacher und häufig analytisch lösbar werden.

¹Vgl. z. B. Bluhm et al. (2010, Kap. 2), Höse (2007), Henking et al. (2006, Kap. 6) oder McNeil et al. (2005, Kap. 8) zu Mischungsmodellen für die Ausfall- und Faktormodellen für die Bonitätsvariablen.

²Empirische Untersuchungen zu dieser Annahme bei Unternehmensausfällen sind z. B. in Kramer & Löffler (2010), Lando & Nielsen (2010) oder Das et al. (2007) zu finden.

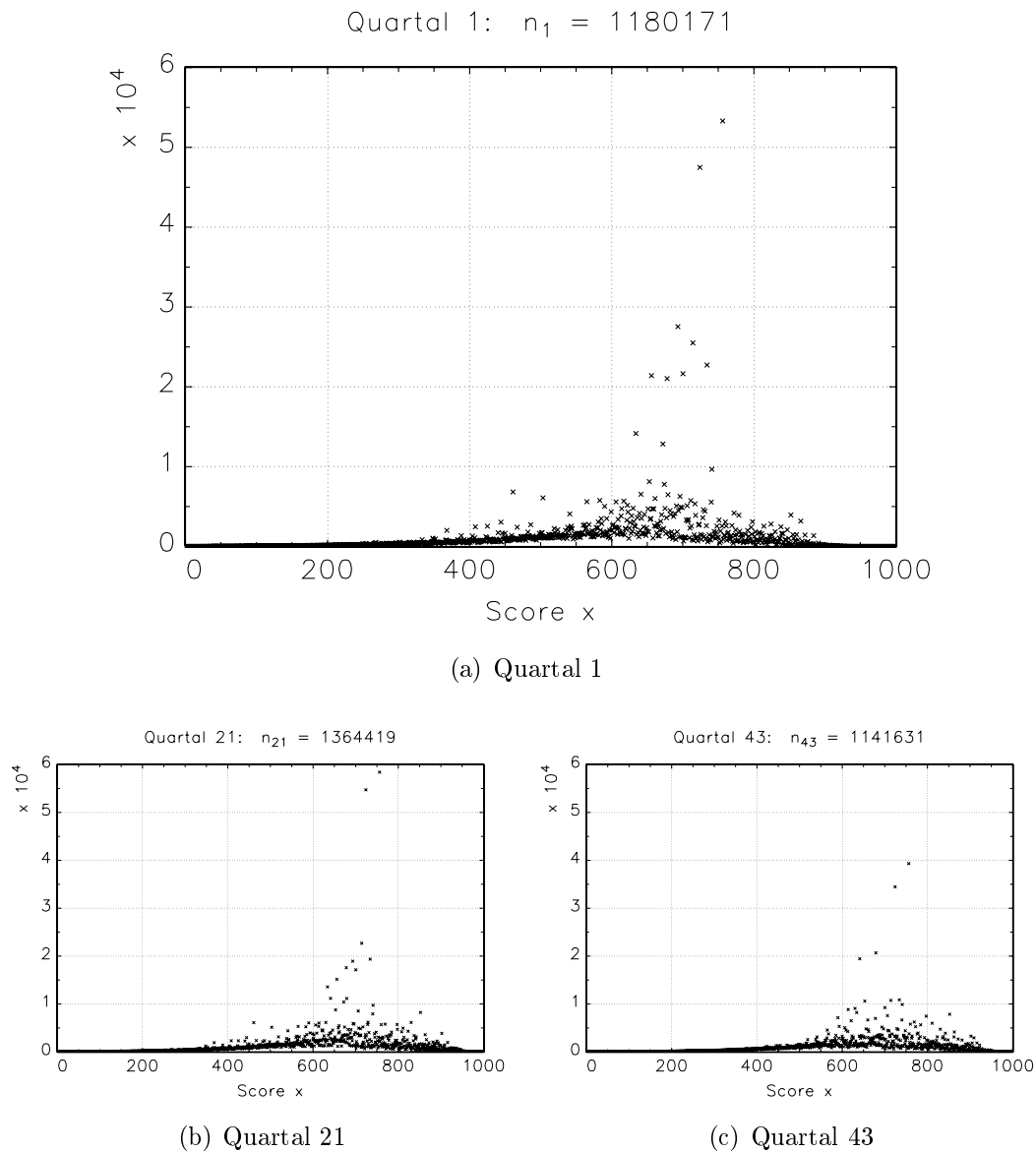


Abbildung 4.2: Absolute Häufigkeit der Bonitätsscores für drei ausgewählte Quartale

4.1.2 Univariate Analysen

Die Bonitätsscores X_{it} sind diskret verteilt, weil nur die natürlichen Zahlen von 1 bis 1000 angenommen werden können, die Score-Verteilung also einen endlichen Träger besitzt. Um uns einen Eindruck über die Verteilung der Scores zu verschaffen, betrachten wir zunächst ihre absoluten Häufigkeiten. Diese sind beispielhaft für die Quartale 1, 21 und 43 in Abbildung 4.2 dargestellt. Dabei bezeichnet n_t jeweils den Stichprobenumfang im relevanten Quartal. Auffällig sind gewisse „Ausreißerscores“ mit sehr großen Besetzungszahlen, bei denen auch die resultierenden empirischen Verteilungsfunktionen, welche in Abbildung 4.3 dargestellt sind, deutlich sichtbare Sprünge haben. Insbesondere gilt dies für die Scores 724 und 756, die in allen betrachteten Quartalen über 34 000 Beobachtungen enthalten. Vermutlich entstehen diese überdurchschnittlich

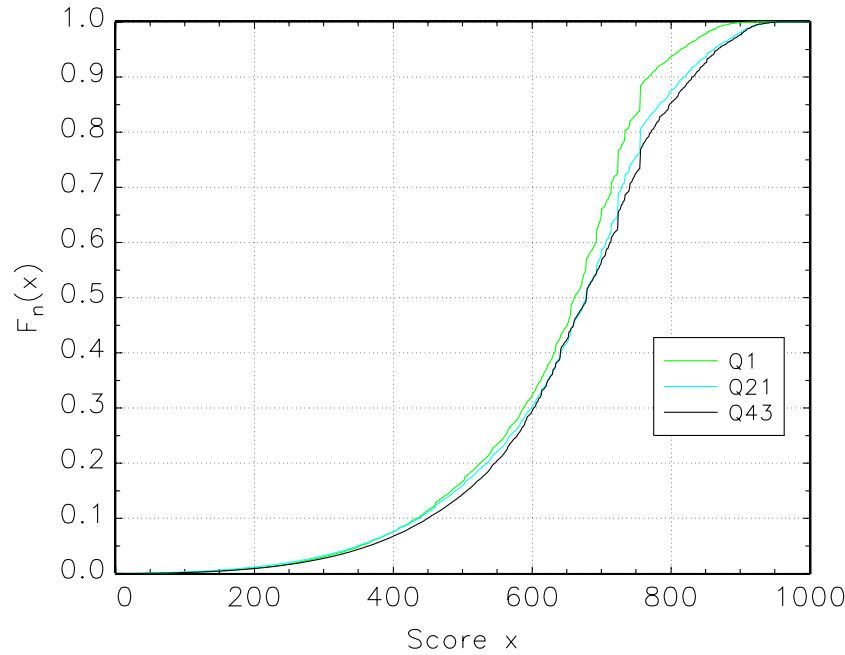


Abbildung 4.3: Empirische Verteilungsfunktion der Bonitätsscores für drei ausgewählte Quartale

hohen Häufigkeiten durch in der Bevölkerung sehr häufig vorkommende Kombinationen von Eingangsvariablen bei der Scoreberechnung. Möglicherweise spielt auch der Einfluss diskreter Input-Variablen eine Rolle.

Wegen der diskreten Verteilung der Scorevariablen sind die für die asymptotischen Eigenschaften der Bruchpunktschätzer hinreichenden lokalen Voraussetzungen an F aus den Sätzen und Folgerungen in Kapitel 3 *nicht* erfüllt. Beispielsweise schließen sich hier Stetigkeit und strenge Monotonie aus. Wir können also nicht sicher sein, dass die Schätzer stark konsistent oder sogar n -konsistent sind oder sich eine fast schwache Konvergenz der Schätzfehlerverteilungen ergibt. Aber ist diese Sicherheit für unser Problem notwendig? Brauchen wir eine möglichst exakte Schätzung von z. B. $\theta = 500.415$? Aus Sicht des Autors lautet die Antwort nein. Zu Referenzzwecken genügt es, die Größenordnung der Trennstelle zu kennen. Ähnliches gilt auch für die später berechneten Konfidenzintervalle. Hier ist besonders die Breite der Konfidenzintervalle von Interesse.

Führt man sich die in Abschnitt 2.1.1 skizzierte Entstehung der Bonitätsscores vor Augen, wird zudem klar, dass eventuell erst bei der Transformation der geschätzten individuellen Ausfallwahrscheinlichkeiten auf die Scores wünschenswerte Eigenschaften der Verteilungsfunktion verloren gehen. Ist z. B. mindestens eine der verwendeten bonitätsrelevanten Eingangsvariablen stetig verteilt, dann sind auch der Rohscore und die Schätzer der individuellen Ausfallwahrscheinlichkeit stetig verteilt. Erst durch die Transformation zu Scores entsteht eine diskrete Skala. Unterstellt man in diesem Fall

die Existenz einer Dichtefunktion für die Bonitätsvariable, kann diese Dichtefunktion geschätzt werden. Die im Folgenden durchgeführte Kerndichteschätzung ist zudem

- ein nützliches Hilfsmittel für die später durchgeführte Kernregression zur Überprüfung der Treppengestalt der Regressionsfunktion κ und
- eine Voraussetzung für die Bestimmung der Konfidenzintervalle und Konfidenzregionen in Abschnitt 4.2.3.

Bei der Kerndichteschätzung wird vorausgesetzt, dass eine i.i.d.-Stichprobe $Z_i \stackrel{\text{i.i.d.}}{\sim} Z$, $i = 1, \dots, n$ zu einer stetigen Zufallsvariable Z mit unbekannter Dichtefunktion f vorliegt.³ Geschätzt wird die Dichtefunktion durch den Kerndichteschätzer⁴

$$\hat{f}_h(x) = \frac{1}{nh} \sum_{i=1}^n \mathcal{K} \left(\frac{x - Z_i}{h} \right), \quad x \in \mathbb{R}. \quad (4.1)$$

Dabei ist $h > 0$ die Bandbreite (*bandwidth, window width*) und \mathcal{K} die Kern- oder Gewichtsfunktion (*kernel, window, weighting function*). Häufig wird die Bandbreite abhängig vom Stichprobenumfang n gewählt und erhält dann den Index n , d. h. $h = h_n$. Es gibt eine Vielzahl von Kernfunktionen,⁵ deren Auswahl bei praktischen Anwendungen jedoch kaum Einfluss auf die Schätzung hat.⁶ Einen viel stärkeren Einfluss auf das Ergebnis hat die Bandbreite, die das Ausmaß der Glättung festlegt.⁷ So führt eine kleine Bandbreite zu einer schwachen Glättung der Daten, was zwar eine geringe Verzerrung, aber eine hohe Varianz des Dichteschätzers nach sich zieht. Umgekehrt führt eine große Bandbreite zu einer starken Glättung der Daten, d. h. zu einer geringen Varianz, aber starken Verzerrung des Dichteschätzers. Für den Ausgleich der Einflüsse von Verzerrung und Varianz geht man zum mittleren quadratischen Fehler oder zum integrierten mittleren quadratischen Fehler über. Die Minimierung der Fehler führt auf eine optimale Bandbreite, die jedoch im Allgemeinen von der unbekannten Dichte f und ihrer – demzufolge ebenfalls unbekannten – zweiten Ableitung f'' abhängig ist. Daher wird in der Praxis für die Festlegung der Bandbreite häufig auf sogenannte Cross-Validation-Techniken zurückgegriffen.⁸

Obwohl die Beschränktheit der Scores gegen ihn spricht, werden wir für unsere Zwecke auf den Normalkern (*Gaussian kernel*) $\mathcal{K}(x) = \varphi(x)$, d. h. die Dichtefunktion der

³Zu Problemen der Kerndichteschätzung, wenn diese Standardvoraussetzungen verletzt sind, vgl. z. B. Hansen (2008, abhängige Daten) oder Zinde-Walsh (2008, nicht existierende Dichtefunktion).

⁴Vgl. Wasserman (2006, S. 132, Definition 6.25), Härdle (1991, S. 45), van der Vaart (1998, S. 342) oder Silverman (1994, S. 15).

⁵Für Beispiele vgl. z. B. Wasserman (2006, S. 55), Härdle (1991, S. 45), Silverman (1994, S. 43) oder Parzen (1962, S. 1068). Teilweise werden dabei gleiche Kernfunktionen unterschiedlich bezeichnet.

⁶Vgl. Wasserman (2006, S. 132), Härdle (1991, S. 78) oder van der Vaart (1998, S. 342).

⁷Vgl. Wasserman (2006, S. 50-54, 134), Härdle (1991, S. 48, 54-60), van der Vaart (1998, S. 342-343).

⁸Vgl. Wasserman (2006, S. 126-127, 135-136), Silverman (1994, S. 48-55) oder Härdle (1991, S. 92-102). In Härdle et al. (2004, S. 79) befinden sich Literaturhinweise zu weiteren Verfahren zur Wahl der Bandbreite.

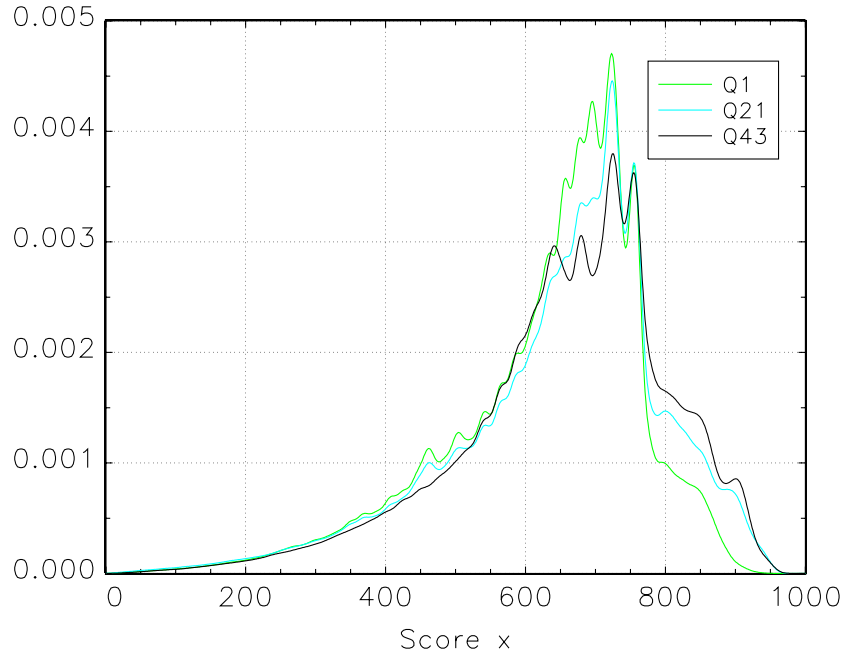


Abbildung 4.4: Kerndichtfunktionen \hat{f}_h der Bonitätsscores für drei ausgewählte Quartale

Standardnormalverteilung, zurückgreifen. Für diese Kernfunktion gibt es nämlich eine einfache Faustregel (*Normal reference rule*), die die Bandbreite h der Kernschätzung festlegt.⁹ In der quartalsbezogenen Betrachtung ergibt sich damit¹⁰

$$h = h_n = 1.06n^{-\frac{1}{5}} \min \left\{ s^*, \frac{\hat{Q}}{1.34} \right\}. \quad (4.2)$$

Dabei bezeichnet $s^* = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2}$ die empirische korrigierte Standardabweichung¹¹ der Scores, $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$ das arithmetische Mittel der Scores und $\hat{Q} = \tilde{x}_{0.75} - \tilde{x}_{0.25}$ den empirischen Quartilabstand. Für die Bandbreite ergeben sich Werte zwischen 7.717 im Quartal 3 und 8.864 im Quartal 39, wenn sie in jedem Quartal gemäß (4.2) berechnet wird. In Abbildung 4.4 sind beispielhaft die Kerndichtfunktionen für die Quartale 1, 21 und 43 dargestellt.

Durch Integrieren kann aus der Kerndichteschätzung ein Schätzwert \hat{F}_{KD} für die Verteilungsfunktion der Scores gewonnen werden:

$$\hat{F}_{\text{KD}}(x) = \int_{-\infty}^x \hat{f}_h(u) du.$$

⁹Vgl. Wasserman (2006, S. 135), Silverman (1994, S. 45-48), Härdle et al. (2004, S. 51-53) oder Härdle (1991, S. 91).

¹⁰Um die Notation übersichtlich zu halten, wird bei den Ausführungen zur Kerndichteschätzung der Zeitindex t unterdrückt.

¹¹Im Gegensatz dazu ist die empirische Standardabweichung s definiert als $s = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2}$.

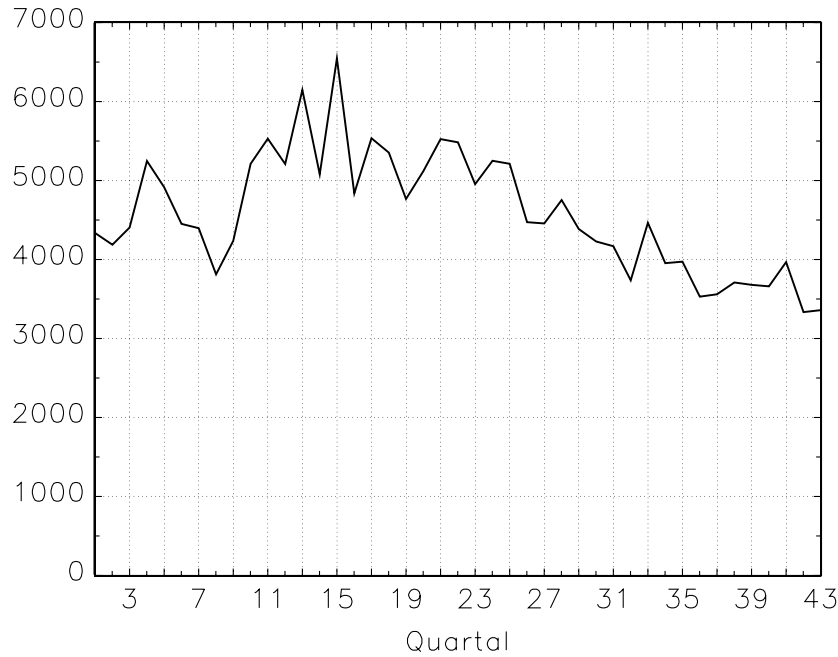


Abbildung 4.5: Anzahl der Ausfälle $\sum_{i=1}^{n_t} y_{it}$ im Zeitablauf

Im vorliegenden Fall gilt

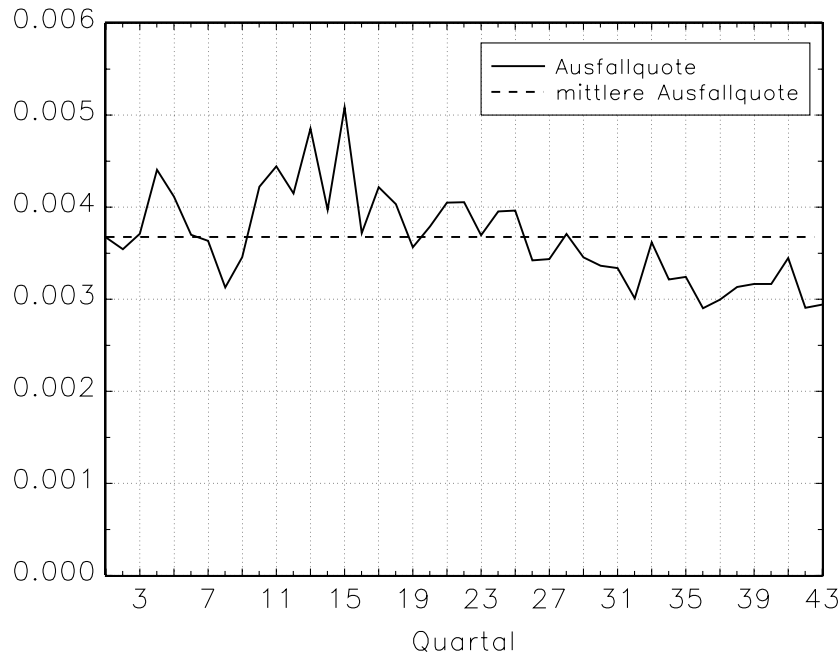
$$\hat{F}_{\text{KD}}(x) = \int_{-\infty}^x \frac{1}{nh} \sum_{i=1}^n \varphi\left(\frac{u-X_i}{h}\right) du = \frac{1}{n} \sum_{i=1}^n \int_{-\infty}^x \frac{1}{h} \varphi\left(\frac{u-X_i}{h}\right) du = \frac{1}{n} \sum_{i=1}^n \Phi\left(\frac{x-X_i}{h}\right).$$

Die Verteilungsfunktion der Standardnormalverteilung Φ ist auf ganz \mathbb{R} streng monoton wachsend und stetig differenzierbar mit $\Phi'(x) = \varphi(x) > 0$. Da φ beschränkt ist, folgt aus der Differenzierbarkeit von Φ nicht nur die Stetigkeit, sondern auch die Lipschitz-Stetigkeit.¹² Wegen Lemma 3.40 sind die Bedingungen in (3.30) und (3.55) vollständig gegeben. Somit erfüllt Φ alle für die asymptotischen Eigenschaften der Bruchpunktschätzer hinreichenden lokalen Voraussetzungen an die Verteilungsfunktion von X aus den Sätzen und Folgerungen in Kapitel 3. Das überträgt sich wegen der Endlichkeit der Summenbildung auch auf die geschätzte Verteilungsfunktion \hat{F}_{KD} .

Nachdem wir uns ausführlich mit der Verteilung der Bonitätsscores befasst haben, interessiert uns auch die Verteilung der Ausfälle. In Abbildung 4.5 ist die Zahl der quartalsweisen Ausfälle $\sum_{i=1}^{n_t} y_{it}$ im zeitlichen Verlauf dargestellt. Die resultierenden quartalsweisen Ausfallquoten $\bar{y}_t = \frac{1}{n_t} \sum_{i=1}^{n_t} y_{it}$ sind in Abbildung 4.6 visualisiert. Darüber hinaus enthält die Abbildung die mittlere Ausfallquote über alle Quartale

$$\bar{y} = \frac{1}{\sum_{t=1}^{43} n_t} \sum_{t=1}^{43} \sum_{i=1}^{n_t} y_{it} \approx 0.0036758.$$

¹²Vgl. Heuser (1994, S. 280, Satz 49.4 in Verbindung mit S. 136 u. S. 212).

Abbildung 4.6: Ausfallquoten \bar{y}_t und \bar{y} im Zeitablauf

Die Ausfallquoten sind sehr klein. Das liegt zum einen an der quartalsweisen Betrachtung und zum anderen daran, dass Kunden aus dem Gesamtbestand der Auskunftei und nicht nur Kreditnachfrager bzw. -nehmer berücksichtigt sind. Im Zeitverlauf sind tendenziell sinkende Ausfallquoten zu beobachten. Dies dürfte auch eine Nachwirkung der alternden Kohorte sein, denn Individuen erreichen mit wachsendem Alter tendenziell eine höhere Bonität. Vergleiche in diesem Zusammenhang auch die Abbildungen 4.3 und 4.4, in denen sich die Bonitätsverbesserung auch widerspiegelt.

4.1.3 Multivariate Analysen

Im Folgenden betrachten wir die Scores der Ausfälle.¹³ Die absoluten Häufigkeiten der einzelnen Ausfallscores sind beispielhaft für Quartal 1 in Abbildung 4.7 dargestellt. Dabei bezeichnet n_D die Anzahl der Ausfälle. Die resultierende empirische Verteilungsfunktion, d. h. die Hit Rate HR , von Quartal 1 ist zusammen mit denen der Quartale 21 und 43 in Abbildung 4.8 dargestellt. Die Kurven verlaufen merklich weiter links oben als die der zugehörigen empirischen Verteilungsfunktionen F_n aller Scores (vergleiche Abbildung 4.3). Kleine Scores sind unter den Ausfällen also deutlich häufiger vertreten als in der Gesamtbetrachtung. Das bestätigen auch die Ergebnisse der Kerndichteschätz-

¹³Dabei und auch im weiteren Verlauf der Arbeit wird der Zeitindex t zur Vereinfachung der Notation unterdrückt. Wird ein spezieller Zeitpunkt betrachtet, lässt er sich aus dem Kontext erschließen.

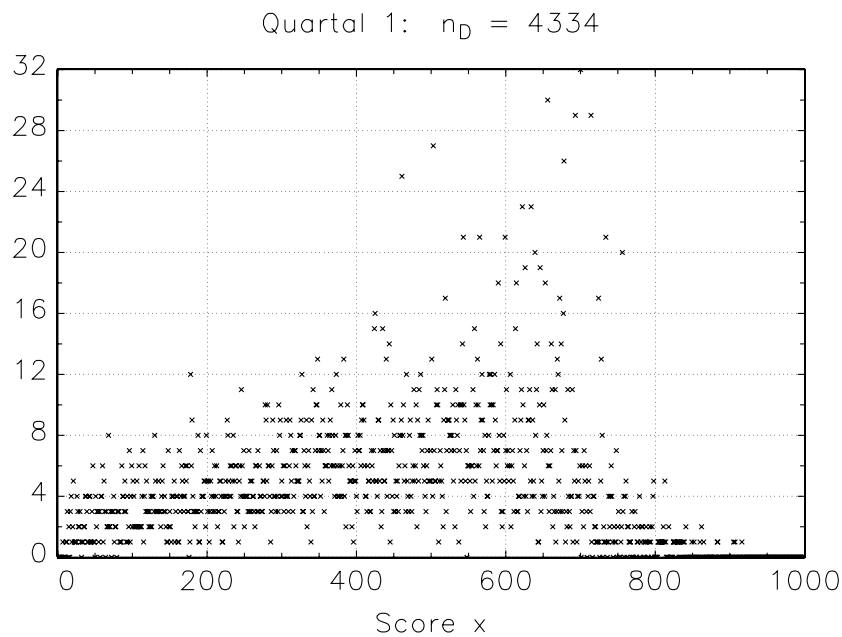


Abbildung 4.7: Absolute Häufigkeit der Bonitätsscores der Ausfälle für ein ausgewähltes Quartal

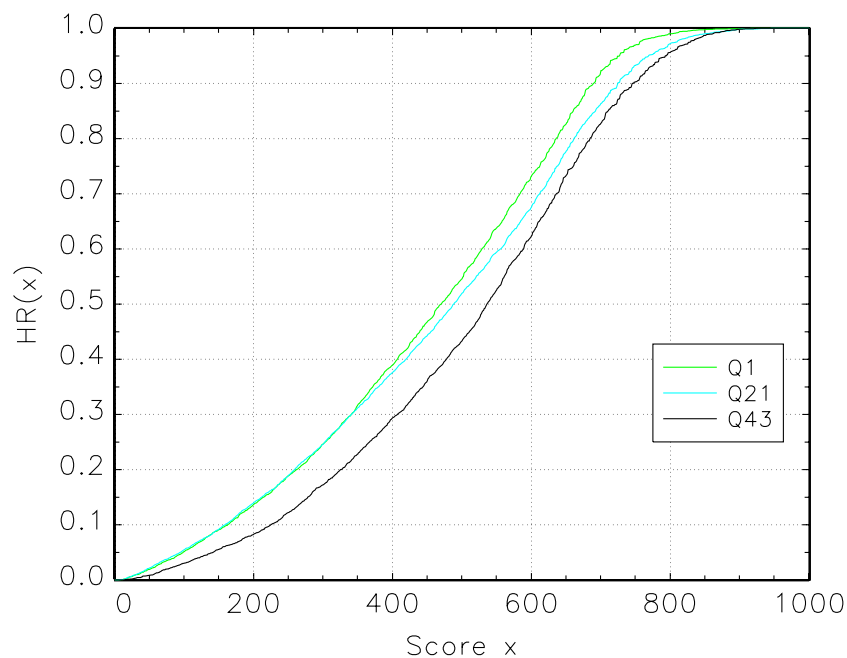


Abbildung 4.8: Empirische Verteilungsfunktion HR der Bonitätsscores der Ausfälle für drei ausgewählte Quartale

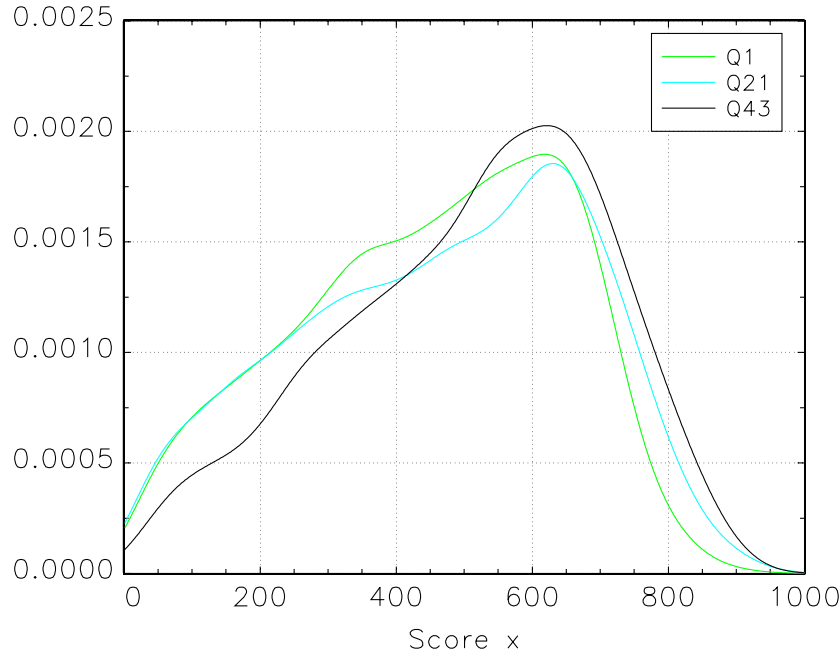


Abbildung 4.9: Kerndichtfunktionen \hat{f}_h^D für die Bonitätsscores der Ausfälle in drei ausgewählten Quartalen

zung. Wird sie für die Scores der Ausfälle durchgeführt, ergibt sich der Schätzer

$$\hat{f}_h^D(x) = \frac{1}{N_D h} \sum_{i=1}^{N_D} \mathcal{K}\left(\frac{x - X_i^D}{h}\right) = \frac{1}{\bar{Y} n h} \sum_{i=1}^n \mathcal{K}\left(\frac{x - X_i}{h}\right) Y_i, \quad (4.3)$$

wobei $N_D = \sum_{i=1}^n Y_i = n\bar{Y}$ die zufällige Anzahl der Ausfälle im entsprechenden Zeitraum und X_i^D den Score des i -ten Ausfalls bezeichnet. Das zweite Gleichheitszeichen ergibt sich maßgeblich wegen $Y_i \in \{0, 1\}$ und der Interpretation von $Y_i = 1$ als Ausfall.

Für die konkreten Schätzwerte müssen noch die Kernfunktion und die Bandbreite festgelegt werden. Als Kernfunktion wird erneut der Normalkern $\mathcal{K}(x) = \varphi(x)$ verwendet. Die Bandbreite wird gemäß der Faustregel in Gleichung (4.2) berechnet. Dabei ergeben sich Werte zwischen 37.592 im Quartal 15 und 43.571 im Quartal 36. Diese Bandbreiten sind unübersehbar größer als die bei der Kerndichteschätzung aller Scores. Das liegt zum einen und vornehmlich am kleineren Stichprobenumfang ($n_D \ll n$), zum anderen aber auch an der größeren Streuung der Ausfallscores. Folge der größeren Bandbreiten sind sehr glatte Kerndichtfunktionen. Vergleiche dazu Abbildung 4.9, in der beispielhaft die Kerndichteschätzwerte \hat{f}_h^D für die Quartale 1, 21 und 43 dargestellt sind. Auf eine gesonderte Betrachtung der Verteilungen der Nichtausfall-Scores und der zugehörigen Verteilungsfunktionen *FAR* (*False Alarm Rates*) wird verzichtet, weil sie sich wegen der geringen Ausfallquoten (vergleiche Abbildung 4.6) und dem funktionalen Zusammenhang aus Lemma 2.1 kaum von den empirischen Verteilungen aller Scores unterscheiden.

Quartal	unbesetzte Scores	
1	1-6, 8, 9	944, 951, 954, 956-1000
21	1, 2, 4	964, 965, 969-1000
43	1-12	960, 965-967, 969-1000

Tabelle 4.1: Unbesetzte Scores in drei ausgewählten Quartalen

Aus den bisherigen Untersuchungen ist bereits ersichtlich geworden, dass die Ausfälle vermehrt bei den kleineren Scores auftreten. Im Folgenden werden wir genauer auf den Zusammenhang von Ausfallverhalten und Score eingehen. Bei über einer Million Beobachtungen in jedem Quartal aber nur 1000 möglichen Scores sind Bindungen zwangsläufig. Als Bindungen werden wiederholte Beobachtungen derselben Merkmalsausprägung bezeichnet. Folglich ist es sinnvoll, die empirische Regression erster Art zu ermitteln. Dazu wird für jeden Merkmalswert der erklärenden Variable X der Mittelwert der zugehörigen Werte der erklärten Variable Y berechnet.¹⁴ In unserem Fall wird also für jeden Score $x = 1, \dots, 1000$ die scorespezifische Ausfallquote

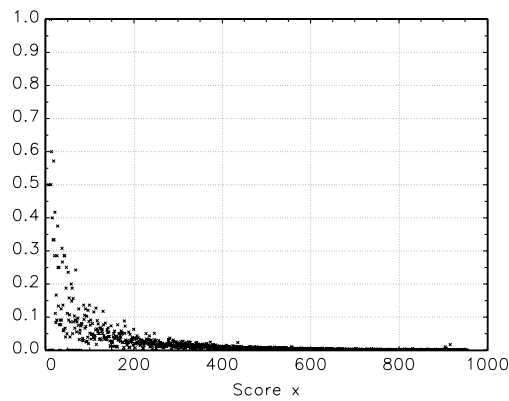
$$\bar{Y}_{[x]} = \frac{\sum_{i=1}^n \mathbb{1}_{\{X_i=x\}} Y_i}{\sum_{i=1}^n \mathbb{1}_{\{X_i=x\}}} \quad (4.4)$$

bestimmt. Es ist möglich, dass (4.4) nicht definiert ist. Dies ist dann der Fall, wenn der entsprechende Score x unbesetzt ist und somit der Nenner auf der rechten Seite null ist. In Tabelle 4.1 sind die unbesetzten Scores für die Quartale 1, 21 und 43 aufgelistet. Es ist zu erkennen, dass nur Scores am Beginn und am Ende der möglichen Skala von 1 bis 1000 unbesetzt sind. Diese Beobachtung gilt tatsächlich für alle Quartale. Die Scores 19 bis 929 sind in allen 43 Quartalen besetzt.

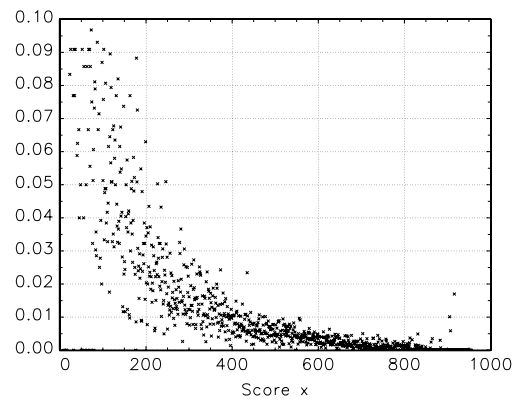
Abbildung 4.10 zeigt die scorespezifischen Ausfallquoten $\bar{Y}_{[x]}$ in den Quartalen 1, 21 und 43. Dabei ist der Schätzwert für die Regressionsfunktion auf der linken Seite vollständig dargestellt, auf der rechten Seite wird die Ordinatenachse jeweils auf das Intervall von 0 bis 0.1 beschränkt. Abgebildet ist also jeweils ein Ausschnitt aus der Grafik auf der linken Seite. Offensichtlich liefert Abbildung 4.10 keinen Beleg für die Treppengestalt der Regressionsfunktion, wie sie im Modell in (3.2) unterstellt wird. Die Grafiken sprechen vielmehr für einen stetigen Verlauf der Regressionsfunktion κ . Üblicherweise wird im Kreditrisiko unterstellt, dass das Ausfallrisiko mit wachsender Bonität exponentiell sinkt.¹⁵ Abbildung 4.11 unterstützt diese These: Die von null ver-

¹⁴Vgl. z. B. Rinne (2008, S. 79) oder Härdle (1993, S. 14).

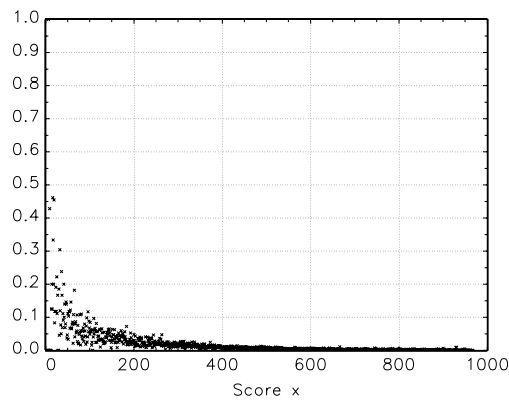
¹⁵Vgl. z. B. Henking et al. (2006, S. 249) oder Bluhm et al. (2010, S. 15).



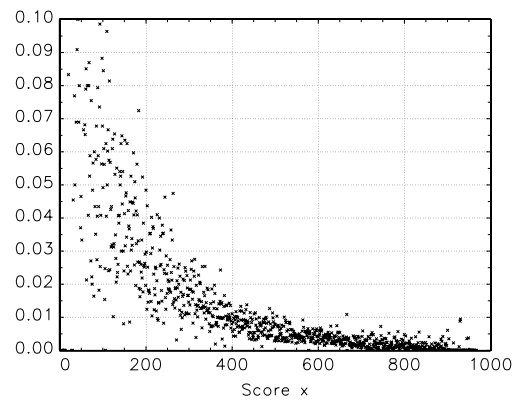
(a) Quartal 1



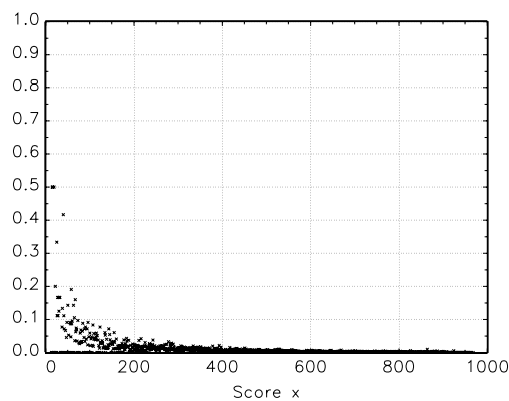
(b) Quartal 1 (Ausschnitt)



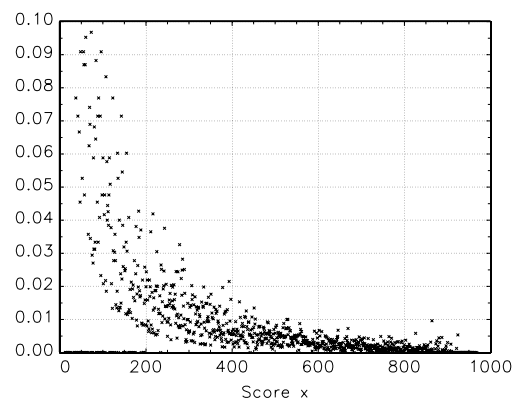
(c) Quartal 21



(d) Quartal 21 (Ausschnitt)



(e) Quartal 43



(f) Quartal 43 (Ausschnitt)

Abbildung 4.10: Empirische Regression 1. Art für drei ausgewählte Quartale

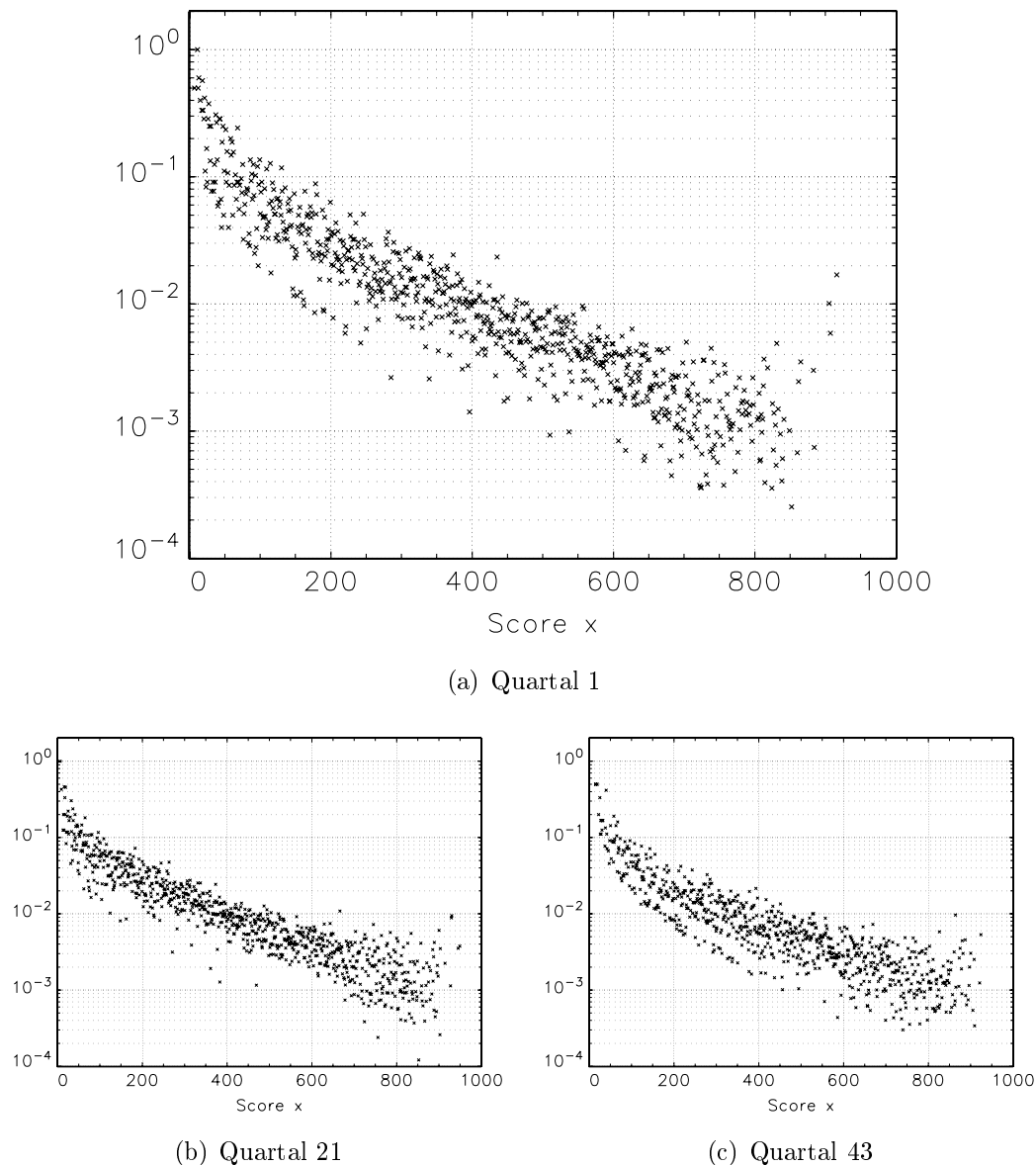


Abbildung 4.11: Empirische Regression 1. Art für drei ausgewählte Quartale, logarithmische Skala

schiedenen¹⁶ scorespezifischen Ausfallquoten verlaufen auf einer logarithmischen Skala annähernd linear.

Um mit den vorhandenen Mitteln einen noch besseren Eindruck vom funktionalen Zusammenhang zwischen Ausfallwahrscheinlichkeit und Bonitätsscore zu erhalten, bietet sich eine nichtparametrische Regression an. Sie ist hier ebenso von Vorteil, da sie nicht schon von vornherein einen streng monotonen Verlauf der Regressionsfunktion unterstellt, wie das z. B. bei der logistischen Regression der Fall wäre. Als Spezialfall der nichtparametrischen Regression wird eine Kernregression mit dem sogenannten

¹⁶Alle Scores mit einer Ausfallquote von null mussten für die Grafik entfernt werden.

Nadaraya-Watson-Schätzer¹⁷

$$\hat{\kappa}_h(x) = \frac{\sum_{i=1}^n \mathcal{K}\left(\frac{x-X_i}{h}\right) Y_i}{\sum_{j=1}^n \mathcal{K}\left(\frac{x-X_j}{h}\right)}$$

durchgeführt. Dieser kann wegen $Y_i \in \{0, 1\}$ sowie (4.1) und (4.3) zu

$$\hat{\kappa}_h(x) = \frac{\bar{Y} \frac{1}{\bar{Y}nh} \sum_{i=1}^n \mathcal{K}\left(\frac{x-X_i}{h}\right) Y_i}{\frac{1}{nh} \sum_{j=1}^n \mathcal{K}\left(\frac{x-X_j}{h}\right)} = \bar{Y} \frac{\hat{f}_h^D(x)}{\hat{f}_h(x)}$$

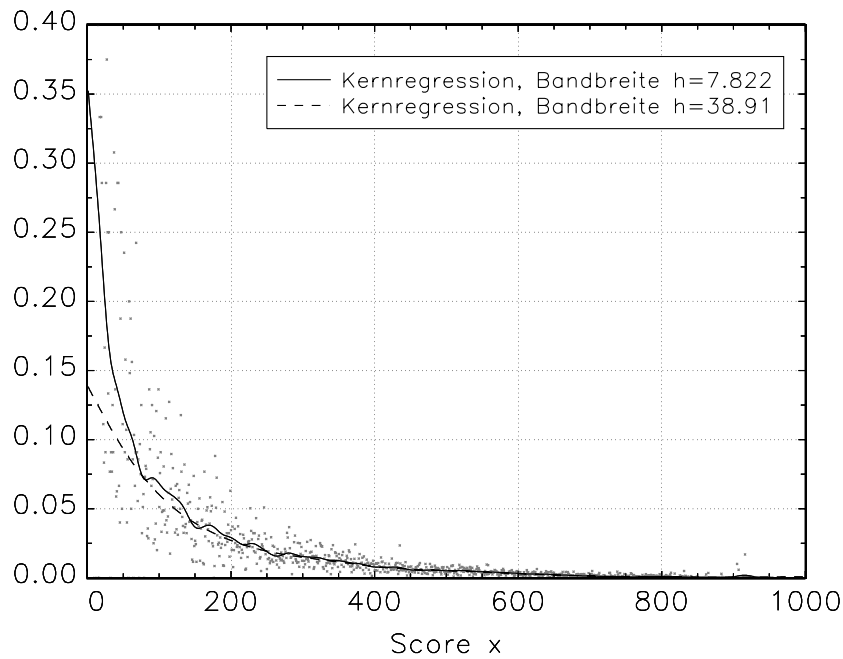
umgeformt werden, wenn $\bar{Y} > 0$ gilt. Somit ist der Nadaraya-Watson-Schätzer in unserem Fall ein mit der Ausfallquote \bar{Y} des entsprechenden Zeitraums skalierter Quotient von Kerndichteschätzern. Im Zähler des Quotienten steht der Kerndichteschätzer für die Dichtefunktion der Scores der Ausfälle. Im Nenner steht der Kerndichteschätzer für die Dichtefunktion aller Scores. Dabei muss jedoch für beide Schätzer dieselbe Bandbreite verwendet werden.

Die für die Quartale 1, 21 und 43 resultierenden Schätzwerte sind in Abbildung 4.12 veranschaulicht. Dabei wurde wie bei der Kerndichteschätzung der Normalkern $\mathcal{K}(x) = \varphi(x)$ verwendet. Es sind zwei Bandbreiten dargestellt. Die kleinere ist diejenige aus der Kernschätzung der Dichtefunktion aller Scores.¹⁸ Die größere Bandbreite stimmt mit der Bandbreite des Kerndichteschätzers für die Dichtefunktion der Scores der Ausfälle überein und liefert einen stärker geglätteten Schätzwert. Wegen der Verwendung eines stetigen Kerns ist auch die Kernregressionsfunktion stetig. Das muss für die wahre Regressionsfunktion nicht gelten. Die Kernregressionsfunktion ist also auch bei sehr großem Stichprobenumfang nur eine Approximation der wahren Regressionsfunktion.¹⁹ Nichtsdestotrotz liegt auch bei den Kernregressionsfunktionen offensichtlich keine Treppengestalt vor. Es sind weder konstante noch annähernd konstante Teilstücke erkennbar. Das ist auch wenig verwunderlich, denn die Grundlage für die Scores ist eine logistische Regression. Diese enthält in ihrem Modell einen stetigen, monotonen Einfluss des Scores auf die Ausfallwahrscheinlichkeit. Wir verwenden die Funktionsgestalt aus (3.2), d. h. die einstufige Treppe, demnach nur als erste Näherung, um die Individuen anhand ihres Scores in solche mit hoher bzw. niedriger Ausfallwahrscheinlichkeit zu trennen. Dieses Modell entspricht zudem der im Kreditrisiko üblichen Forderung nach einer einheitlichen Ausfallwahrscheinlichkeit innerhalb jeder Ratingklasse, wobei anzumerken ist, dass dann in der Regel mehr als zwei Ratingklassen betrachtet werden.

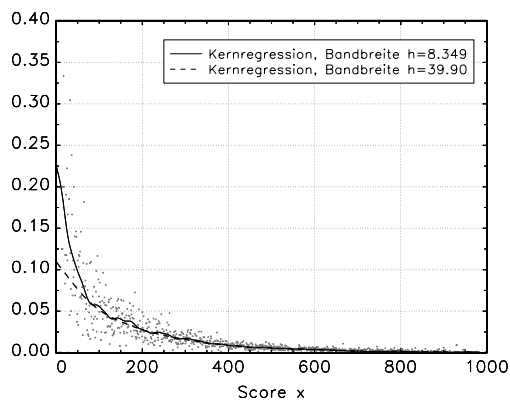
¹⁷Vgl. Wasserman (2006, S. 71), Härdle (1991, S. 123-127), Härdle (1993, S. 24-25) oder Härdle et al. (2004, S. 88-90). In allen genannten Arbeiten werden weitere nichtparametrische Regressionsmethoden vorgestellt.

¹⁸Diese Bandbreite erfüllt die in Li & Racine (2007, S. 66) genannte Faustregel.

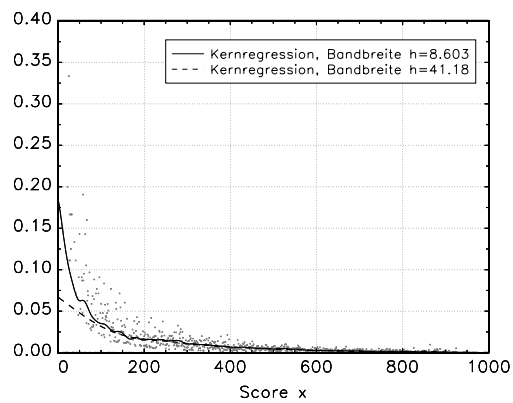
¹⁹Vgl. Härdle (1993, S. 15).



(a) Quartal 1



(b) Quartal 21



(c) Quartal 43

Abbildung 4.12: Kernregression für drei ausgewählte Quartale (verschiedene Bandbreiten) im Vergleich zur empirischen Regression 1. Art

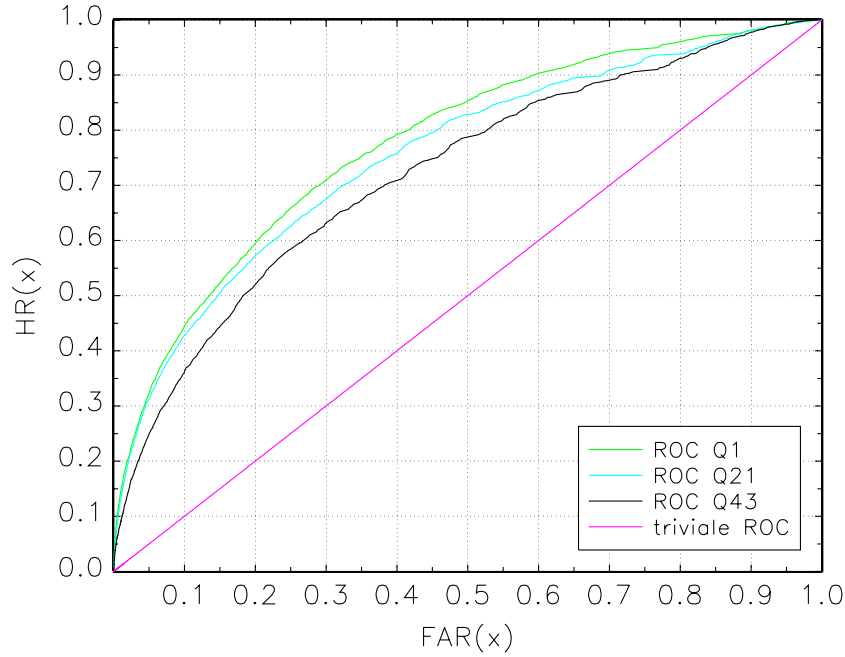


Abbildung 4.13: ROC-Kurven für die Quartale 1, 21 und 43

Zum Abschluss von Abschnitt 4.1 widmen wir uns kurz der Güte des Scoringmodells der Auskunftei. Grundlage für die Gütemessung bilden die im Abschnitt 2.1.2 vorgestellten Konzepte. Zunächst ist in Abbildung 4.13 beispielhaft für die Quartale 1, 21 und 43 jeweils die ROC dargestellt. Die ROC-Kurven aller anderen Quartale besitzen qualitativ einen ähnlichen Verlauf. Auf ihre Darstellung wird ebenso verzichtet wie auf die Darstellung der CAP-Kurven, die sich wegen der geringen Ausfallquoten nur kaum sichtbar von den jeweiligen ROC-Kurven unterscheiden. Basierend auf der ROC kann das ROC-gap als skalares Gütemaß berechnet werden. Es ist gemäß (2.8) das Maximum der Funktion d^{ROC} , welche die Differenz zwischen der empirischen Verteilungsfunktion der Ausfälle (Hit Rate, HR) und der empirischen Verteilungsfunktion der Nichtausfälle (False Alarm Rate, FAR) abbildet. Die Funktion d^{ROC} wird in Abbildung 4.14 beispielhaft für die drei Quartale 1, 21 und 43 visualisiert. Sie kann gemäß Folgerung 3.5 als Grundlage für die Schätzung des Bruchpunkts zwischen den zwei Ratingklassen verwendet werden. Die aus der Abstandsfunktion d^{ROC} resultierenden quartalsweisen ROC-gaps sind in Abbildung 4.15 dargestellt. Weil in allen Quartalen die Bedingung (2.11) für die Identität²⁰ von ROC-gap und KS-Statistik erfüllt ist, gibt Abbildung 4.15 gleichermaßen letzteres Gütemaß an. Die Ausprägungen dieser Maßzahlen liegen im Rahmen der in der Praxis üblichen Werte.²¹ Im Zeitverlauf zeigen sie eine sinkende Tendenz. Daraus sollte aber nicht ohne Weiteres geschlossen werden, dass

²⁰Wir schließen uns dem Teil der Literatur an, in dem die Prüfgröße des Kolmogoroff-Smirnoff-Tests der maximale Abstand der betrachteten empirischen Verteilungsfunktionen ist.

²¹Vgl. Mays (2004a, S. 111).

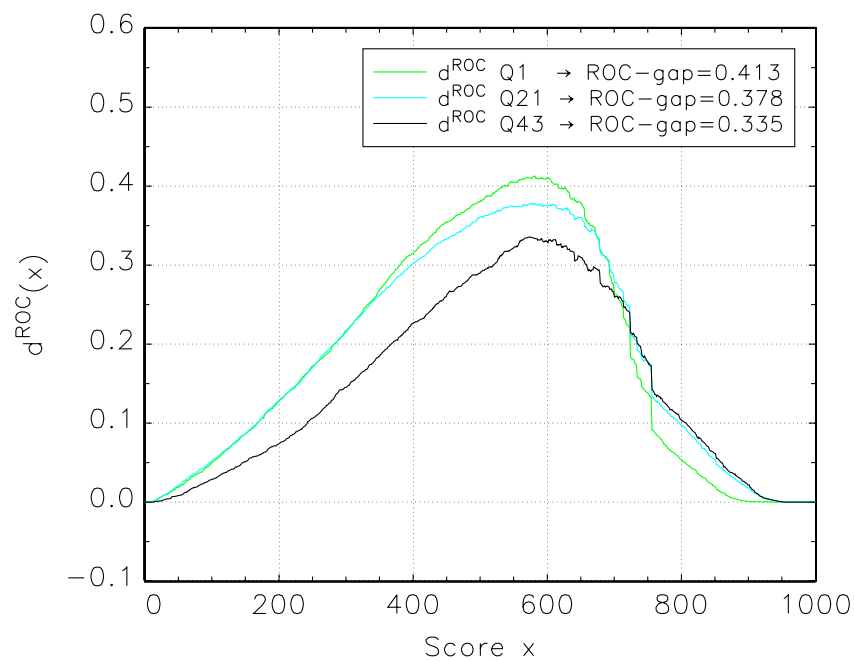


Abbildung 4.14: Funktion d^{ROC} für die Quartale 1, 21 und 43

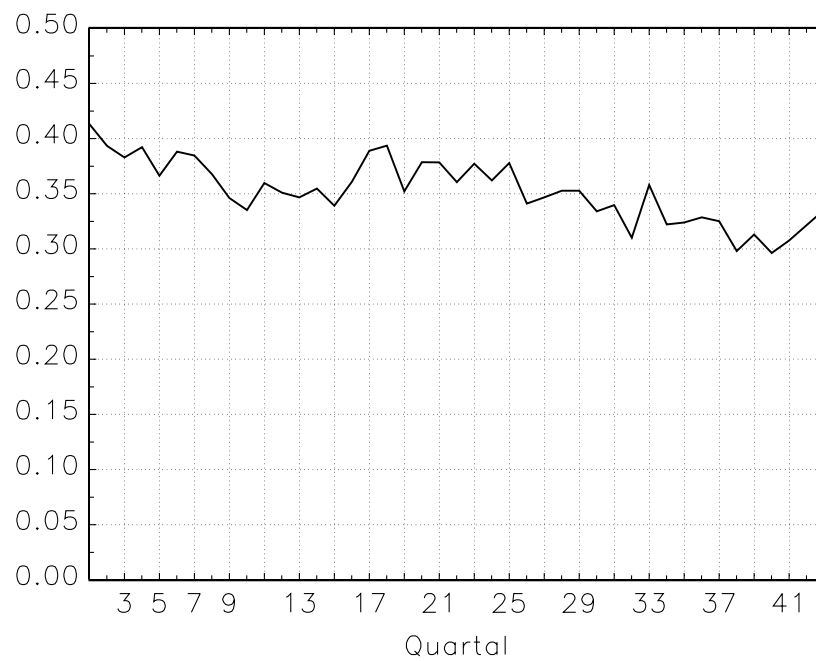


Abbildung 4.15: ROC-gap (=KS-Statistik) im Zeitablauf

die Güte des Scoringsystems der Auskunftfeil abnimmt, denn aus der ROC oder dem CAP abgeleitete Maßzahlen hängen entscheidend von der Struktur und der Qualität, d. h. den Ausfallwahrscheinlichkeiten, des zugrunde liegenden Portfolios ab.²² Schon wegen der Kohortenbetrachtung muss hier von Änderungen ausgegangen werden.

4.2 Bruchpunktschätzung für einzelne Quartale

Im folgenden Abschnitt werden die Methoden der Bruchpunktschätzung aus Kapitel 3 im soeben beschriebenen Datensatz angewendet. Die Untersuchung erfolgt an dieser Stelle quartalsweise, also nur im zeitlichen Querschnitt.²³ Es werden dabei die Fälle mit

1. vorgegebenen, bekannten Risikoniveaus und
2. unbekannten Risikoniveaus

unterschieden.

Im ersten Fall wird der Maximum-Likelihood-Schätzer aus Abschnitt 3.3 verwendet. Die resultierenden ML-Schätzwerte werden mit den in Abschnitt 2.2.1 vorgestellten *benchmark breakpoints* nach Fernandes (2005, S. 39) und Anderson (2007, S. 422) verglichen. Die ebenso in Abschnitt 2.2.1 erwähnte naive Klasseneinteilung mit den Mittelpunkten der Masterskala kann nicht angewendet werden, da im vorliegenden Datensatz keine Schätzwerte $\hat{\pi}_i$ für die individuellen Ausfallwahrscheinlichkeiten verfügbar sind.

Im zweiten Fall der unbekannten Risikoniveaus stehen der Dempfle-Stute-Schätzer aus Abschnitt 3.2 und der Zwei-Schritt-Plug-In-Schätzer aus Abschnitt 3.4 im Mittelpunkt der Betrachtung. Referenzwerte liefern die zwei naiven Ansätze aus Abschnitt 2.2.1. Dabei werden der Mittelpunkt des Scorebereichs, hier also die Zahl 500, und der Median der empirischen Scoreverteilung als Trennstelle zwischen den Klassen herangezogen. Sowohl die Clusteranalyse als auch die Verfahren aus Krink et al. (2008, S. 2278) werden nicht angewendet, da für diese Methoden die individuellen PD-Schätzwerte $\hat{\pi}_i$ notwendig wären.²⁴ Als Vergleichswert eines fortgeschrittenen Optimierungsansatzes soll die Trennstelle dienen, die die Calinski-Harabasz-Statistik (CH-Statistik) aus Abschnitt 2.2.2 maximiert.

Zum Vergleich der Güte der angesprochenen Methoden wird unter anderem eben-diese CH-Statistik verwendet. Dies erfolgt auch mit dem Ziel, den Einsatz der CH-Statistik als Gütekriterium generell zu überprüfen. Als weitere Qualitätsmerkmale wird

²²Vgl. Blochwitz et al. (2004), Blümke (2011) und speziell für die KS-Statistik Mays (2004b).

²³Die periodenübergreifende Auswertung erfolgt in Abschnitt 4.3.

²⁴Für den Ansatz in (2.19) wären zudem Annahmen über Preise/Zinssätze von Nöten. Dies wäre hier wenig sinnvoll, da nicht jede gescorte Person tatsächlich einen Kredit haben muss.

auf die Zeitstabilität der geschätzten Trennstellen und die Besetzung der resultierenden Klassen eingegangen. Abschließend werden für den Fall bekannter Risikoniveaus Konfidenzintervalle für den Bruchpunkt und im Fall unbekannter Risikoniveaus simultane Konfidenzregionen für den Bruchpunkt und die Risikoniveaus berechnet. Auch in diesem Abschnitt der Arbeit wird der Zeitindex t zur Vereinfachung der Notation unterdrückt. Wird ein spezieller Zeitpunkt betrachtet, lässt er sich aus dem Kontext erschließen.

4.2.1 Maximum-Likelihood-Schätzer

Im Folgenden werden die Risikoniveaus, d. h. in unserem Kontext die Ausfallwahrscheinlichkeiten, als bekannt vorausgesetzt. Speziell gelte

$$a = 0.010 = 1.0\% \quad \text{und} \quad b = 0.001 = 0.1\%, \quad (4.5)$$

solange nichts anderes angegeben ist. Im Vergleich zu den Eingangsparametern der Simulationsstudie in Ferger & Klotsche (2009, S. 120-121), hier beschrieben in Abschnitt 3.5, ist der Abstand der Risikoniveaus äußerst gering.

Der Maximum-Likelihood-Schätzwert für den unbekannten Bruchpunkt θ ist die kleinste Maximalstelle des empirischen Prozesses S_n^* gemäß (3.27), d. h.

$$\theta_n^* := \operatorname{argmax}_{x \in \mathbb{R}} S_n^*(x).$$

Für die drei Quartale 1, 21 und 43 ist der Funktionsverlauf des zugehörigen empirischen Prozesses S_n^* in Abbildung 4.16 dargestellt. Die kleinste Maximalstelle und damit der Bruchpunktschätzwert ist der Score 580 in Quartal 1, der Score 608 in Quartal 21 und der Score 561 in Quartal 43.

Hätte man andere Eingangsparameter als in (4.5) gewählt, wäre der Funktionsverlauf des empirischen Prozesses S_n^* von der Form her ähnlich. Vergleiche dazu Abbildung 4.17, in der die Funktionsverläufe für vier verschiedene Parameterkonstellationen beispielhaft für Quartal 1 visualisiert sind. In der Abbildung ist zu erkennen, dass der Prozess zunächst tendenziell wächst, um ab einem gewissen Punkt wieder zu fallen. Der Übergang von Wachstum zu Abnahme erfolgt dabei nicht abrupt, d. h. der Gipfel der Funktion S_n^* ist nicht spitz, sondern eher flach. Das spricht für einen kontinuierlichen Verlauf der Regressionsfunktion und gegen die Vorstellung, dass die Regressionsfunktion wie im vorgeschlagenen Modell stückweise konstant ist.²⁵

²⁵Das ist jedoch keine Kritik an dem Modell, welches bekanntlich nur eine vereinfachte Abbildung der Realität ist und sein soll. Ferner entspricht das Modell genau der im Kreditrisiko häufig getätigten Annahme, dass alle Kreditnehmer einer Ratingklasse eine einheitliche Ausfallwahrscheinlichkeit besitzen.

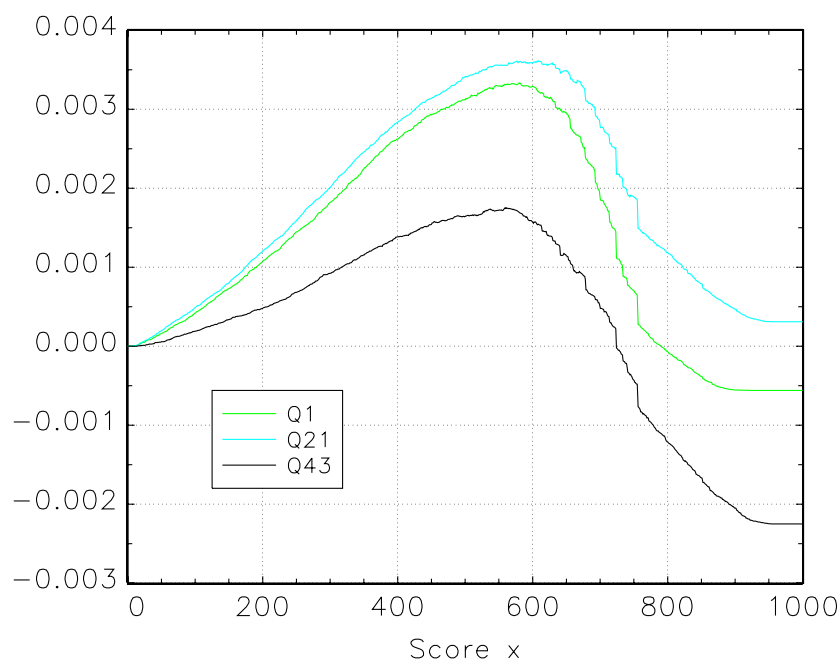


Abbildung 4.16: Funktionsverlauf des empirischen Prozesses S_n^* (Maximum-Likelihood) für drei ausgewählte Quartale bei bekannten Risikoniveaus $a = 0.01$ und $b = 0.001$: Maximalstellen bei 580 (Q1), 608 (Q21) und 561 (Q43)

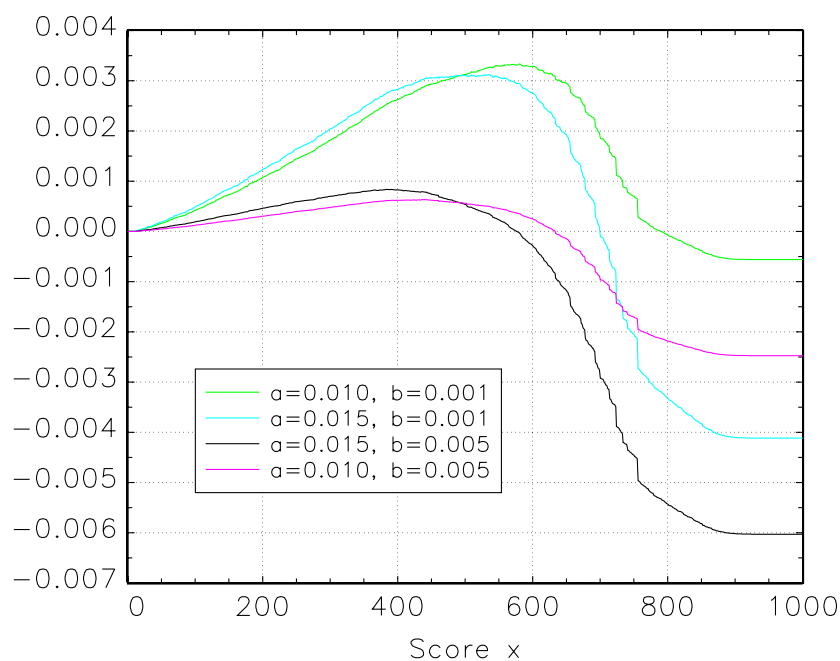


Abbildung 4.17: Funktionsverlauf des empirischen Prozesses S_n^* (Maximum-Likelihood) im Quartal 1 für vier ausgewählte Parameterkonstellationen

Bevor wir die Auswertungen für alle Quartale des Untersuchungszeitraums vorstellen, werden im Folgenden vier konkurrierende Ansätze dargestellt. Dabei handelt es sich um die in Abschnitt 2.2.1 einführend besprochenen Methoden nach Fernandes (2005, S. 39) und Anderson (2007, S. 422, Gleichung 20.2) sowie um zwei darauf basierende Ansätze. Um eine gemeinsame Basis für alle Verfahren zu schaffen, wird zunächst die Notation an die aus Kapitel 3 angepasst.

Das Ziel sind $K = 2$ Ratingklassen mit den vorgegebenen Referenzausfallwahrscheinlichkeiten $y_1^B = a$ und $y_2^B = b$. Die Ratingklassen sollen anhand der Scores der n beobachteten Individuen eingeteilt werden. Ist x die Trennstelle zwischen den Ratingklassen, dann besteht die Ratingklasse 1 aus allen Individuen mit einem Score kleiner oder gleich x und Ratingklasse 2 besteht aus allen Individuen mit einem Score echt größer als x . In Mengenschreibweise erhält man

- Ratingklasse 1: $\{i = 1, \dots, n | X_i \leq x\}$,
- Ratingklasse 2: $\{i = 1, \dots, n | X_i > x\}$.

Die Anzahl der Individuen in den Ratingklassen ist von der Trennstelle x und von der Stichprobe abhängig und daher zufällig. Es gilt²⁶

$$N_1 = N_1(x) = |\{i = 1, \dots, n | X_i \leq x\}| = \sum_{i=1}^n \mathbb{1}_{\{X_i \leq x\}} = nF_n(x),$$

$$N_2 = N_2(x) = n - N_1(x) = n(1 - F_n(x)).$$

Als Ausfallquoten für die zwei durch x getrennten Ratingklassen ergeben sich²⁷

$$\bar{Y}_1 = \bar{Y}_1(x) = \frac{\sum_{j=1}^{N_1} Y_{1j}}{N_1} = \frac{\sum_{i=1}^n \mathbb{1}_{\{X_i \leq x\}} Y_i}{\sum_{i=1}^n \mathbb{1}_{\{X_i \leq x\}}} = \frac{H_n(x)}{F_n(x)} = a_n^*(x)$$

und analog

$$\bar{Y}_2 = \bar{Y}_2(x) = \frac{\sum_{j=1}^{N_2} Y_{2j}}{N_2} = \frac{\sum_{i=1}^n \mathbb{1}_{\{X_i > x\}} Y_i}{\sum_{i=1}^n \mathbb{1}_{\{X_i > x\}}} = \frac{\bar{Y} - H_n(x)}{1 - F_n(x)} = b_n^*(x).$$

Dabei bezeichnet Y_{kj} die j -te Beobachtung der Ausfallvariable in der Ratingklasse k .

Mit dieser Notation ergibt sich für die Methodik aus Fernandes (2005, S. 39) der nachstehende Schätzer für die Trennstelle zwischen den zwei Ratingklassen:

$$\hat{\theta}_F := \operatorname{argmin}_{x \in \mathbb{R}} \{(a - a_n^*(x))^2 + (b - b_n^*(x))^2\}.$$

²⁶Zur Definition von F_n vgl. Gleichung (2.1), S. 10.

²⁷Vgl. (3.46) und (3.47) auf S. 69.

Vergleiche dazu Gleichung (2.13) auf S. 22 dieser Arbeit. Die sogenannten *benchmark breakpoints* aus Anderson (2007, S. 422, Gleichung 20.2) führen auf den Schätzer

$$\hat{\theta}_A := \operatorname{argmin}_{x \in \mathbb{R}} \left\{ nF_n(x) \left(\ln \left(\frac{1-a}{a} \right) - \ln \left(\frac{1-a_n^*(x)}{a_n^*(x)} \right) \right)^2 + n(1-F_n(x)) \left(\ln \left(\frac{1-b}{b} \right) - \ln \left(\frac{1-b_n^*(x)}{b_n^*(x)} \right) \right)^2 \right\}.$$

Vergleiche dazu Gleichung (2.14) auf S. 22. In Bemerkung 2.4 wurde im Zusammenhang mit diesen beiden Schätzern die Frage aufgeworfen, inwiefern das Einbeziehen oder Weglassen der Klassenbesetzungen N_1 und N_2 die Ergebnisse der Verfahren beeinflusst. Zur Beantwortung dieser Frage werden im Folgenden auch die modifizierten Schätzer

$$\hat{\theta}_{F^*} := \operatorname{argmin}_{x \in \mathbb{R}} \{ nF_n(x) (a - a_n^*(x))^2 + n(1 - F_n(x)) (b - b_n^*(x))^2 \}$$

und

$$\hat{\theta}_{A^*} := \operatorname{argmin}_{x \in \mathbb{R}} \left\{ \left(\ln \left(\frac{1-a}{a} \right) - \ln \left(\frac{1-a_n^*(x)}{a_n^*(x)} \right) \right)^2 + \left(\ln \left(\frac{1-b}{b} \right) - \ln \left(\frac{1-b_n^*(x)}{b_n^*(x)} \right) \right)^2 \right\}$$

betrachtet.

Werden die Schätzwerte gemäß den fünf vorgestellten Methoden für alle 43 Quartale des Datensatzes bestimmt, ergibt sich die Darstellung in Abbildung 4.18. Dabei zeigt sich eine typische Reihenfolge der Schätzwerte. In jedem einzelnen Quartal gilt

$$\hat{\theta}_F \leq \hat{\theta}_{F^*} \leq \theta_n^* \leq \min\{\hat{\theta}_A, \hat{\theta}_{A^*}\},$$

wobei das Gleichheitszeichen jeweils in mindestens einem Quartal gilt. Die beiden Schätzwerte nach Anderson verhalten sich nicht in allen Quartalen einheitlich, d. h.

$$\hat{\theta}_A \not\leq \hat{\theta}_{A^*} \quad \text{und} \quad \hat{\theta}_A \not\geq \hat{\theta}_{A^*}.$$

Die Schätzwerte $\hat{\theta}_{F^*}$ und $\hat{\theta}_{A^*}$ der modifizierten Verfahren zeigen im zeitlichen Verlauf ein sehr ähnliches Verhalten wie die zugehörigen Originale $\hat{\theta}_F$ bzw. $\hat{\theta}_A$. Des Weiteren ist festzustellen, dass sich die ML-Schätzwerte θ_n^* in den Grenzen bewegen, die von den praxiserprobten Ansätzen vorgegeben werden. Ferner zeigt sich, auch beim Blick auf die Lage- und Streuungsmaßzahlen in Tabelle 4.2, dass alle Schätzwerte eine große Spannweite aufweisen. Die ML-Schätzwerte sind im Vergleich zu den anderen Metho-

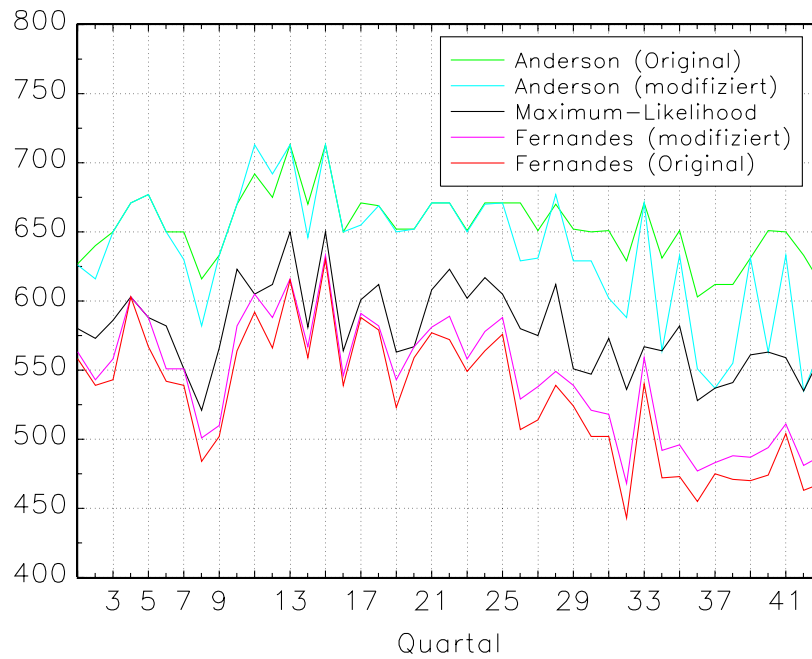


Abbildung 4.18: Bruchpunktschätzwerte im Zeitablauf, bekannte Risikoniveaus $a = 0.01$ und $b = 0.001$

	Fernandes (Original) $\hat{\theta}_F$	Fernandes (modifiziert) $\hat{\theta}_{F^*}$	Maximum- Likelihood θ_n^*	Anderson (modifiziert) $\hat{\theta}_{A^*}$	Anderson (Original) $\hat{\theta}_A$
arithm. Mittel	530.81	544.19	579.19	635.26	653.65
Median	539	549	575	646	651
Minimum	443	468	521	535	603
in Quartal:	32	32	8	42	36
Maximum	630	633	650	713	713
in Quartal:	15	15	13, 15	11, 13, 15	13, 15
Spannweite	187	165	129	178	110
Standardabw.	46.80	42.73	30.77	46.18	24.43

Tabelle 4.2: Lage- und Streuungsmaßzahlen zu den Bruchpunktschätzwerten bei bekannten Risikoniveaus $a = 0.01$ und $b = 0.001$

den jedoch recht zeitstabil. Eine kleinere Spannweite und Standardabweichung weisen nur die Schätzwerte $\hat{\theta}_A$ auf. Die mit den weiteren Ansätzen gewonnenen Schätzwerte streuen dagegen viel stärker. Generell schwanken die Schätzwerte über die Zeit in einem starken Gleichlauf. Die entsprechenden empirischen Korrelationen sind jeweils größer als 0.82. Sie können der Korrelationsmatrix in Tabelle 4.3 entnommen werden.

	Fernandes (Original) $\hat{\theta}_F$	Fernandes (modifiziert) $\hat{\theta}_{F^*}$	Maximum- Likelihood θ_n^*	Anderson (modifiziert) $\hat{\theta}_{A^*}$	Anderson (Original) $\hat{\theta}_A$	Ausfall- quote \bar{y}
$\hat{\theta}_F$	1	0.992	0.839	0.873	0.835	0.961
$\hat{\theta}_{F^*}$	0.992	1	0.849	0.893	0.860	0.961
θ_n^*	0.839	0.849	1	0.824	0.848	0.879
$\hat{\theta}_{A^*}$	0.873	0.893	0.824	1	0.875	0.885
$\hat{\theta}_A$	0.835	0.860	0.848	0.875	1	0.893
\bar{y}	0.961	0.961	0.879	0.885	0.893	1

Tabelle 4.3: Empirische Korrelationsmatrix der Bruchpunktschätzwerte und der Ausfallquote, bekannte Risikoniveaus $a = 0.01$ und $b = 0.001$

Bemerkenswert sind ebenso die hohen Korrelationen der Bruchpunktschätzwerte und der quartalsweisen Ausfallquote (vgl. Tabelle 4.3).

Zum Vergleich der fünf untersuchten Optionen zur Ratingklasseneinteilung wird im Folgenden auch die Calinski-Harabasz-Statistik (CH-Statistik) aus Gleichung (2.20), S. 27 herangezogen. Dabei wird hier nicht die Klassenanzahl als Argument verwendet, denn es ist stets $K = 2$, sondern die Trennstelle x zwischen den Klassen. Mit (2.21) und den oben eingeführten Schreibweisen gilt

$$\begin{aligned}
 CH(x) &= \frac{\frac{1}{K-1} \sum_{k=1}^K N_k(x) (\bar{Y}_k(x) - \bar{Y})^2}{\frac{1}{n-K} \sum_{k=1}^K \sum_{j=1}^{N_k(x)} (Y_{kj} - \bar{Y}_k(x))^2} = \frac{\sum_{k=1}^2 N_k(x) (\bar{Y}_k(x) - \bar{Y})^2}{\frac{1}{n-2} \sum_{k=1}^2 N_k(x) \bar{Y}_k(x) (1 - \bar{Y}_k(x))} \\
 &= \frac{nF_n(x) (a_n^*(x) - \bar{Y})^2 + n(1 - F_n(x)) (b_n^*(x) - \bar{Y})^2}{\frac{1}{n-2} \left(nF_n(x) a_n^*(x) (1 - a_n^*(x)) + n(1 - F_n(x)) b_n^*(x) (1 - b_n^*(x)) \right)}.
 \end{aligned} \tag{4.6}$$

Abbildung 4.19 zeigt im Zeitablauf die Werte der CH-Statistik, die sich ergeben, wenn die verschiedenen Bruchpunktschätzwerte eingesetzt werden. Tabelle 4.4 enthält die zugehörigen Lage- und Streuungsmaßzahlen. In jedem einzelnen Quartal gilt

$$CH(\hat{\theta}_F) \geq CH(\hat{\theta}_{F^*}) \geq CH(\theta_n^*) \geq \max\{CH(\hat{\theta}_A), CH(\hat{\theta}_{A^*})\}.$$

Würde die CH-Statistik als alleiniges Gütekriterium zwischen den fünf vorgestellten Optionen verwendet werden, dann müsste die Wahl also auf die Methode nach Fernandes fallen. Sie erhält in allen Quartalen den jeweils größten Wert. Im weiteren Verlauf

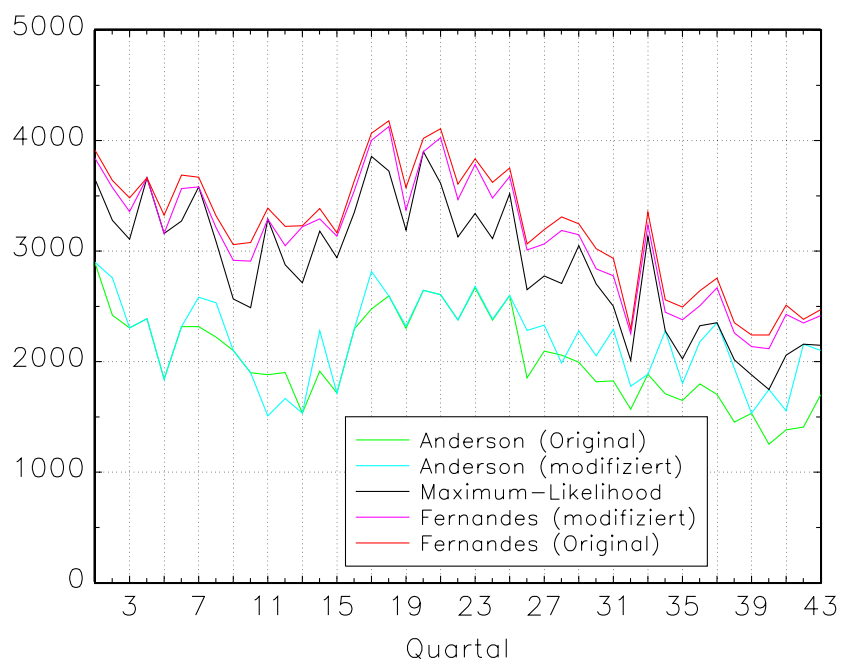


Abbildung 4.19: Werte der Calinski-Harabasz-Statistik für die Bruchpunktschätzwerte im Zeitablauf, bekannte Risikoniveaus $a = 0.01$ und $b = 0.001$

	Fernandes (Original) $CH(\hat{\theta}_F)$	Fernandes (modifiziert) $CH(\hat{\theta}_{F*})$	Maximum- Likelihood $CH(\theta_n^*)$	Anderson (modifiziert) $CH(\hat{\theta}_{A*})$	Anderson (Original) $CH(\hat{\theta}_A)$
arithm. Mittel	3225.0	3124.0	2885.7	2190.0	2029.3
Median	3308.1	3186.0	3047.7	2279.7	1912.6
Minimum	2240.3	2116.2	1748.3	1509.6	1252.9
in Quartal:	40	40	40	11	40
Maximum	4178.7	4126.0	3899.0	2897.0	2886.2
in Quartal:	18	18	20	1	1
Spannweite	1938.4	2009.8	2150.6	1387.4	1633.3
Standardabw.	545.0	549.0	582.0	369.6	402.0

Tabelle 4.4: Lage- und Streuungsmaßzahlen zu den Werten der Calinski-Harabasz-Statistik, wenn die Klassen mit Hilfe der Bruchpunktschätzwerte bei bekannten Risikoniveaus $a = 0.01$ und $b = 0.001$ eingeteilt werden

	Fernandes (Original) $F_n(\hat{\theta}_F)$	Fernandes (modifiziert) $F_n(\hat{\theta}_{F*})$	Maximum- Likelihood $F_n(\theta_n^*)$	Anderson (modifiziert) $F_n(\hat{\theta}_{A*})$	Anderson (Original) $F_n(\hat{\theta}_A)$
arithm. Mittel	0.204	0.221	0.272	0.399	0.438
Median	0.214	0.225	0.257	0.411	0.428
Minimum	0.103	0.124	0.183	0.184	0.299
in Quartal:	32	32	36	42	36
Maximum	0.380	0.386	0.431	0.641	0.637
in Quartal:	15	15	15	11	13
Spannweite	0.276	0.262	0.249	0.457	0.337
Standardabw.	0.067	0.066	0.059	0.116	0.073

Tabelle 4.5: Lage- und Streuungsmaßzahlen zum Anteil der Ratingklasse 1 an allen Beobachtungen, wenn die Klassen mit Hilfe der Bruchpunktschätzwerte bei bekannten Risikoniveaus $a = 0.01$ und $b = 0.001$ eingeteilt werden

der Arbeit werden wir aber feststellen, dass ein Vergleich der Bruchpunktschätzer mit der CH-Statistik wenig sinnvoll ist.

Abschließend werfen wir noch einen Blick auf die Besetzung der resultierenden Ratingklassen. Genauer gesagt betrachten wir den Anteil der Beobachtungen in den beiden Ratingklassen an der Gesamtzahl aller Beobachtungen. Wird beispielsweise der ML-Schätzwert θ_n^* als Trennstelle zwischen den Ratingklassen verwendet, dann beträgt der Anteil der Ratingklasse 1, d. h. der Klasse, in der die niedrigeren Bonitätsscores zusammengefasst sind (vgl. S. 104), gerade $F_n(\theta_n^*) = \frac{N_1(\theta_n^*)}{n}$. Für die weiteren Methoden erfolgt die Berechnung analog. Der komplementäre Anteil der Ratingklasse 2 ergibt sich dann, hier am Beispiel des ML-Schätzwerts, als $1 - F_n(\theta_n^*)$. Insofern genügt es, die Anteile von Ratingklasse 1 anzugeben. Wegen der Monotonie der Verteilungsfunktion gilt generell: Je größer die Trennstelle ist, desto größer ist der Anteil von Ratingklasse 1. Ohne auf alle Details aus Tabelle 4.5 einzugehen, sei angemerkt, dass der Ratingklasse 1 bei allen Methoden mindestens 10.3 % und höchstens 64.1 % der Beobachtungen zugeordnet werden. Der durchschnittliche Anteil von Klasse 1 beträgt je nach Methode zwischen 20.4 % und 43.8 %. Hinsichtlich der Zeitstabilität der Klassenbesetzung liefert die Maximum-Likelihood-Methode die besten Ergebnisse.

4.2.2 Dempfle-Stute- und Plug-In-Schätzer

In diesem Abschnitt sind neben dem Bruchpunkt θ auch die Risikoniveaus a und b unbekannt. Für diesen Fall bieten sich der Dempfle-Stute- und der Zwei-Schritt-Plug-In-Schätzer an, um die gemeinsame Grenze der Ratingklassen zu schätzen. Beide benötigen die Kenntnis der Risikoniveaus nicht. Darüber hinaus existieren weitere Optionen für die Bestimmung der Trennstelle zwischen den Ratingklassen. Wie eingangs angekündigt, liefern die zwei naiven Ansätze aus Abschnitt 2.2.1 erste Referenzwerte:

- Der Mittelpunkt des Scorebereichs, hier also die Zahl 500, trennt die zwei Ratingklassen. Der entsprechende Schätzer sei

$$\hat{\theta}_C := 500.$$

- Der Median der empirischen Scoreverteilung wird als Trennstelle verwendet. Der resultierende Schätzer ist

$$\hat{\theta}_M := \tilde{X}_{0.5} = F_n^{\leftarrow}(0.5).$$

Den dritten Vergleichswert liefert die Calinski-Harabasz-Statistik (CH-Statistik):

- Die optimale Trennstelle im Sinne des CH-Kriteriums ist derjenige Score, der die CH-Statistik maximiert. Der entsprechende Schätzer ist

$$\hat{\theta}_{CH} := \operatorname{argmax}_{x \in \mathbb{R}} CH(x)$$

mit CH gemäß (4.6).

Da die Gesamtstreuung $\sum_{k=1}^K \sum_{j=1}^{N_k} (Y_{kj} - \bar{Y})^2$ der Stichprobe festliegt und die Streuungszerlegung

$$\sum_{k=1}^K \sum_{j=1}^{N_k} (Y_{kj} - \bar{Y})^2 = \sum_{k=1}^K (\bar{Y}_k - \bar{Y})^2 N_k + \sum_{k=1}^K \sum_{j=1}^{N_k} (Y_{kj} - \bar{Y}_k)^2$$

gilt,²⁸ führt die Maximierung der Interklassenstreuung $\sum_{k=1}^K (\bar{Y}_k - \bar{Y})^2 N_k$ zur selben Trennstelle wie die Minimierung der Innerklassenstreuung $\sum_{k=1}^K \sum_{j=1}^{N_k} (Y_{kj} - \bar{Y}_k)^2$. Folglich genügt statt der Maximierung der CH-Statistik auch die einfache (d. h. nicht si-

²⁸Vgl. dazu und im Folgenden auch Bacher et al. (2010, S. 299) oder Everitt et al. (2011, S. 114).

multane) Maximierung des Zählers oder die einfache Minimierung des Nenners der CH-Statistik. Im Zweiklassenfall gilt also

$$\hat{\theta}_{\text{CH}} = \operatorname{argmax}_{x \in \mathbb{R}} \sum_{k=1}^2 (\bar{Y}_k(x) - \bar{Y})^2 N_k(x) = \operatorname{argmin}_{x \in \mathbb{R}} \frac{1}{n-2} \sum_{k=1}^2 \sum_{j=1}^{N_k(x)} (Y_{kj} - \bar{Y}_k(x))^2. \quad (4.7)$$

Bemerkung 4.1 (CH-Schätzer basiert auf geschätzten Regressionsfehlern)

Betrachten wir noch einmal die Größen auf der rechten Seite von Gleichung (4.7): Die Y_{kj} sind die beobachteten Ausfallvariablen, die $\bar{Y}_k(x)$ sind die von der Trennstelle x abhängigen Schätzer für die Niveaus der einstufigen Regressionsfunktion. Ihre Differenz ist der geschätzte Regressionsfehler. Damit ist gezeigt, dass die CH-Methode auf den geschätzten Regressionsfehlern $\hat{\epsilon}_i$ beruht und daher gemäß Dempfle & Stute (2002, S. 234) weniger gut geeignet ist, um den Bruchpunkt θ zu schätzen.

Dempfle-Stute-Schätzer

Auch wenn die Risikonivaus a und b unbekannt sind, können wir ohne Zweifel den Zusammenhang $a > b$ unterstellen, da im vorliegenden Scoringssystem ein kleiner Score für eine geringe Bonität und damit eine große Ausfallwahrscheinlichkeit steht. Umgekehrt bedeutet ein großer Score hohe Bonität und damit eine kleine Ausfallwahrscheinlichkeit. Demzufolge kann neben dem zweiseitigen Dempfle-Stute-Schätzer

$$\bar{\theta}_n := \operatorname{argmax}_{x \in \mathbb{R}} |S_n(x)|,$$

der keine Voraussetzungen an die Relation von a und b stellt, auch der einseitige, d. h.

$$\hat{\theta}_n := \operatorname{argmax}_{x \in \mathbb{R}} S_n(x),$$

verwendet werden.

Der beiden Schätzern zugrunde liegende empirische Prozess S_n gemäß (3.14) ist in Abbildung 4.20 beispielhaft für die drei Quartale 1, 21 und 43 dargestellt. In allen anderen Quartalen hat er qualitativ denselben Verlauf (erst Wachstum, dann Abnahme), so dass auf eine Wiedergabe verzichtet wurde. Der Übergang von Wachstum zu Abnahme erfolgt dabei nicht abrupt, d. h. der Gipfel der Funktion S_n ist nicht spitz, sondern eher flach. Das spricht für einen kontinuierlichen Verlauf der Regressionsfunktion und gegen die Vorstellung, dass die Regressionsfunktion wie im vorgeschlagenen Modell stückweise konstant ist.²⁹

²⁹Das ist jedoch keine Kritik an dem Modell, welches bekanntlich nur eine vereinfachte Abbildung der Realität ist und sein soll. Ferner entspricht das Modell genau der im Kreditrisiko häufig getätig-

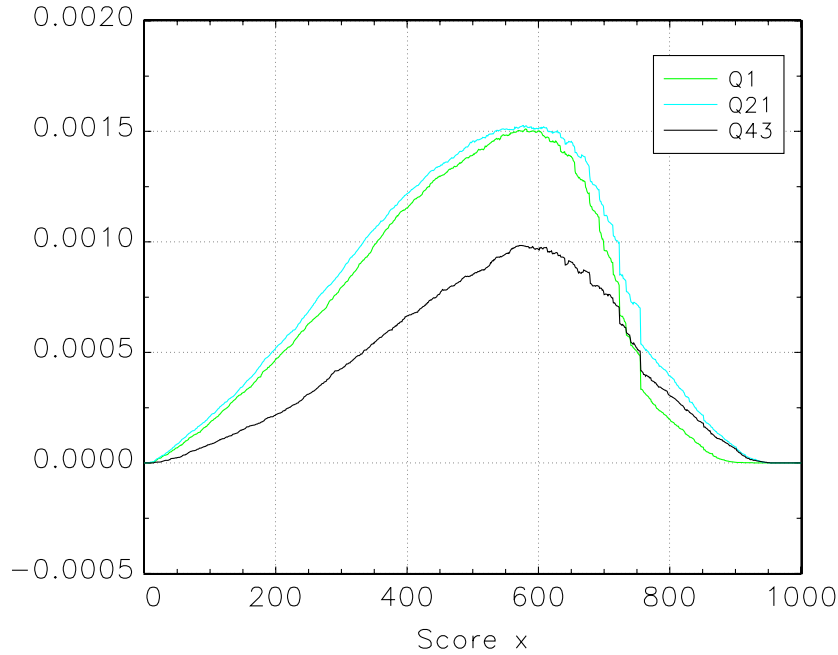


Abbildung 4.20: Funktionsverlauf des empirischen Prozesses S_n (Dempfle-Stute) für drei ausgewählte Quartale: Maximalstellen bei 580 (Q1), 576 (Q21) und 573 (Q43)

In den extremen, schwach besetzten Scorebereichen nahe 0 und nahe 1000 ist der empirische Prozess S_n in einigen Quartalen geringfügig negativ,³⁰ so dass S_n und $|S_n|$ im Allgemeinen nicht gleich sind. In jedem Quartal gilt aber $\max_{x \in \mathbb{R}} S_n(x) > \max_{x \in \mathbb{R}} -S_n(x)$. Daher stimmen die ein- und zweiseitigen Bruchpunktschätzwerte gemäß (3.22) in jedem Quartal überein:

$$\hat{\theta}_n = \bar{\theta}_n.$$

Speziell ergeben sich der Score 580 in Quartal 1, der Score 576 in Quartal 21 und der Score 573 in Quartal 43 als kleinste Maximalstelle von S_n und damit als Dempfle-Stute-Schätzwert für den Bruchpunkt θ . Die Zeitreihe aller Dempfle-Stute-Schätzwerte ist in Abbildung 4.21 dargestellt.

Plug-In-Schätzer

Basierend auf den ermittelten Dempfle-Stute-Schätzwerten können nun Schätzwerte für die unbekannten Risikoniveaus a und b berechnet werden. Speziell betrachten wir

ten Annahme, dass alle Kreditnehmer einer Ratingklasse eine einheitliche Ausfallwahrscheinlichkeit besitzen.

³⁰Dies gilt z. B. auch für die Quartale 21 und 43. Für alle Scores x zwischen 16 und 921 ist $S_n(x)$ in allen Quartalen größer oder gleich null.

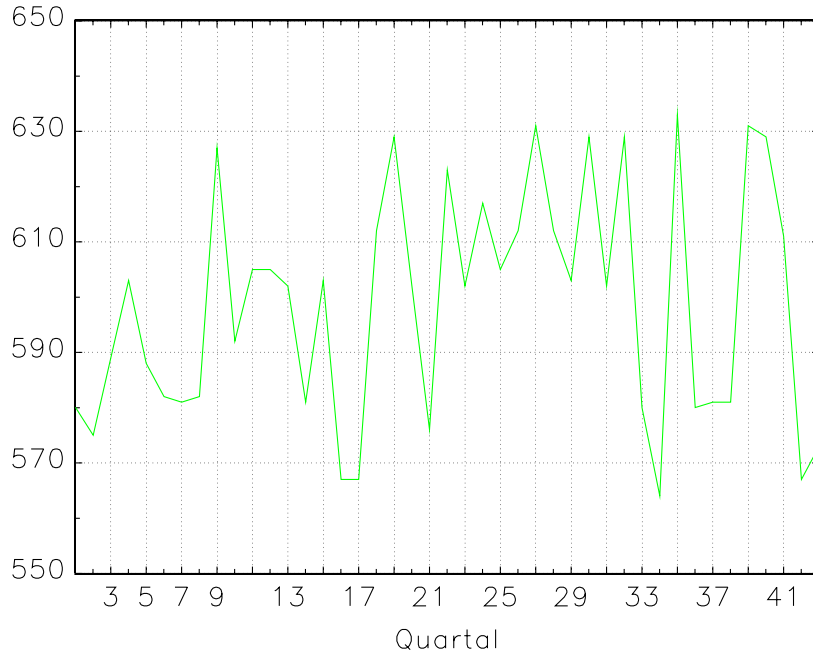


Abbildung 4.21: Bruchpunktschätzwerte nach Dempfle & Stute im Zeitablauf

die Schätzwerte, die aus den Plug-In-Schätzern gemäß (3.51), S. 70 resultieren:

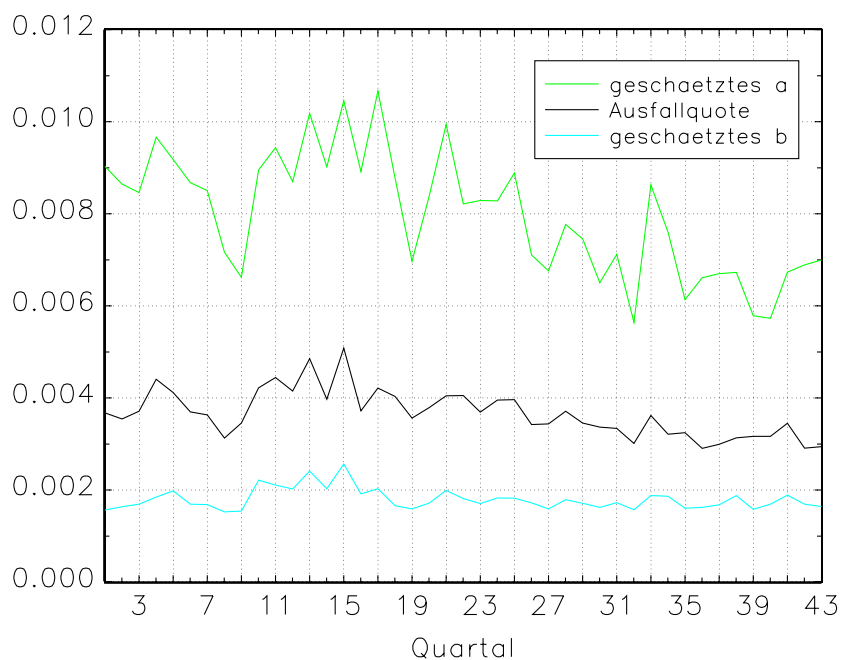
$$\bar{a}_n^* = a_n^*(\bar{\theta}_n) \quad \text{und} \quad \bar{b}_n^* = b_n^*(\bar{\theta}_n). \quad (4.8)$$

Inhaltlich handelt es sich dann um Ausfallquoten für die als schlecht klassifizierten Kreditnehmer (Ratingklasse 1, Score $x \leq \bar{\theta}_n$) und die als gut klassifizierten Kreditnehmer (Ratingklasse 2, Score $x > \bar{\theta}_n$). Da diese Ausfallquoten nur gewisse Teile des Scorebereiches abdecken, werden sie im Folgenden auch als Teilausfallquoten bezeichnet. In Abbildung 4.22 sind die beiden Teilausfallquoten im Vergleich zur jeweiligen Gesamtausfallquote für alle Quartale dargestellt. Die zugehörigen Lage- und Streuungsmaßzahlen sind in Tabelle 4.6 zusammengefasst. In ihr sind ebenso die empirischen Korrelationen zwischen den verschiedenen Ausfallquoten enthalten. Deutlich wird der starke Gleichlauf der (Teil-)Ausfallquoten und die geringe Differenz von \bar{a}_n^* und \bar{b}_n^* , welche im Mittel nur knapp über 0.006 liegt.

Mit Hilfe der Plug-In-Schätzwerte für die Risikoniveaus können nun auch die Substitutionsschätzwerte für den Bruchpunkt θ bestimmt werden. Man erhält sie als die jeweils kleinste Maximalstelle des empirischen Prozesses \bar{S}_n^* gemäß (3.54), d. h.

$$\bar{\theta}_n^* = \operatorname{argmax}_{x \in \mathbb{R}} \bar{S}_n^*(x).$$

Der Funktionsverlauf von \bar{S}_n^* ist in Abbildung 4.23 beispielhaft für die Quartale 1, 21 und 43 visualisiert. Auch hier zeigt sich wieder ein flacher Gipfel der Funktion. Für die

Abbildung 4.22: Zeitreihen der (Teil-)Ausfallquoten \bar{a}_n^* , \bar{y} und \bar{b}_n^*

	Teilausfallquote \bar{a}_n^*	Ausfallquote \bar{y}	Teilausfallquote \bar{b}_n^*
arithm. Mittel	0.00797	0.00366	0.00180
Median	0.00828	0.00363	0.00171
Minimum	0.00564	0.00290	0.00153
in Quartal:	32	36	8
Maximum	0.01067	0.00508	0.00256
in Quartal:	17	15	15
Spannweite	0.00503	0.00218	0.00104
Standardabw.	0.00131	0.00050	0.00022
Korrelationen			
\bar{a}_n^*	1	0.869	0.700
\bar{y}	0.869	1	0.797
\bar{b}_n^*	0.700	0.797	1

Tabelle 4.6: Lage- und Streuungsmaßzahlen sowie empirische Korrelationen zu den (Teil-)Ausfallquoten \bar{a}_n^* , \bar{y} und \bar{b}_n^*

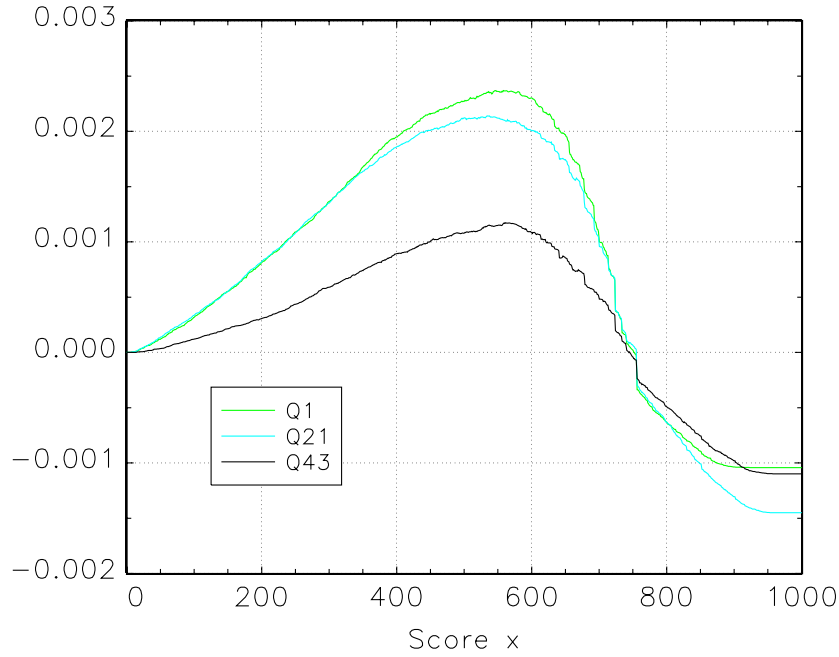


Abbildung 4.23: Funktionsverlauf des empirischen Prozesses \bar{S}_n^* (Plug-In) für drei ausgewählte Quartale: Maximalstellen bei 558 (Q1), 537 (Q21) und 561 (Q43)

drei Quartale von Interesse ergeben sich die Schätzwerte 558, 537 bzw. 561. Schon mit diesen drei Werten wird klar, dass die Plug-In-Schätzwerte im Allgemeinen nicht mit den Dempfle-Stute-Schätzwerten übereinstimmen.

Vergleich der Bruchpunktschätzwerte

Die Zeitreihe aller Plug-In-Schätzwerte für die Trennstelle zwischen den Ratingklassen ist zusammen mit den Zeitreihen der vier alternativen Möglichkeiten in Abbildung 4.24 dargestellt. Dabei zeigt sich eine typische Reihenfolge der Schätzwerte. In jedem einzelnen Quartal gilt

$$\hat{\theta}_{CH} < \hat{\theta}_C = 500 < \bar{\theta}_n^* \leq \bar{\theta}_n = \hat{\theta}_n < \hat{\theta}_M.$$

Die Unterschiede zwischen den Schätzwerten sind dabei zum Teil enorm. Selbst die Differenz zwischen den Dempfle-Stute-Schätzwerten und den Plug-In-Schätzwerten kann bis zu 63 Scorepunkte betragen (Quartal 19). Sie kann aber auch gleich null sein, denn in den Quartalen 9, 12, 34 und 39 stimmen die Schätzwerte $\hat{\theta}_n$ und $\bar{\theta}_n^*$ überein. Im Mittel beträgt ihre Differenz 27.60. Diese doch recht große Differenz scheint bemerkenswert, insbesondere vor dem Hintergrund des Stichprobenumfangs von mehr als einer Million Beobachtungen. Wären alle entsprechenden Annahmen erfüllt, würde man aufgrund der Konsistenz der Schätzer erwarten, dass die Schätzwerte bei einem dermaßen großen Stichprobenumfang viel näher beieinander liegen.

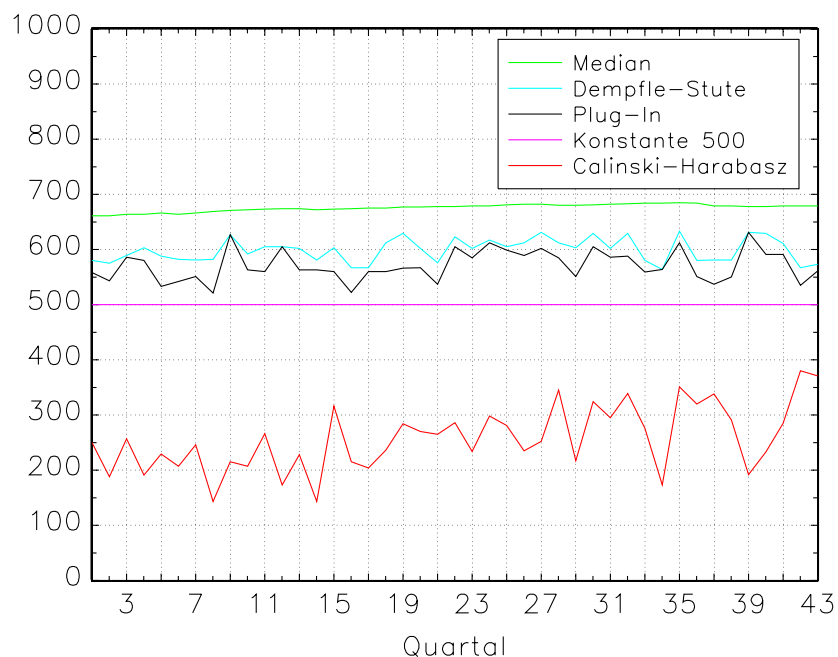


Abbildung 4.24: Bruchpunktschätzwerte im Zeitablauf, unbekannte Risikoniveaus

	Calinski-Harabasz $\hat{\theta}_{CH}$	Konstante $\hat{\theta}_C$	Plug-In $\bar{\theta}_n^*$	Dempfle-Stute $\hat{\theta}_n = \bar{\theta}_n$	Median $\hat{\theta}_M$
arithm. Mittel	256.93	500	571.07	598.67	675.70
Median	252	500	563	602	678
Minimum	143	500	521	564	661
in Quartal:	8, 14	alle	8	34	1, 2
Maximum	380	500	631	633	685
in Quartal:	42	alle	39	35	35
Spannweite	237	0	110	69	24
Standardabw.	58.68	0	27.30	20.82	6.49

Tabelle 4.7: Lage- und Streuungsmaßzahlen zu den Bruchpunktschätzwerten bei unbekannten Risikoniveaus

Einen vertiefenden Blick über die Verteilung der Schätzwerte aller betrachteten Methoden erlauben die Lage- und Streuungsmaßzahlen in Tabelle 4.7. Dabei zeigt sich, dass die beiden naiven Varianten Konstante und Median – zum Teil trivialerweise – äußerst zeitstabile Trennstellen liefern. Die Streuung der nichtparametrischen Bruchpunktschätzwerte $\hat{\theta}_n$ und $\bar{\theta}_n^*$ ist, auch im Vergleich zu den Maximum-Likelihood-Schätzwerten aus Abschnitt 4.2.1, recht klein. Einzig das CH-Kriterium fällt deutlich

	Calinski-Harabasz		Plug-In	Dempfle-Stute	Median	Ausfallquote
	$\hat{\theta}_{\text{CH}}$	$CH(\hat{\theta}_{\text{CH}})$	$\bar{\theta}_n^*$	$\hat{\theta}_n = \bar{\theta}_n$	$\hat{\theta}_{\text{M}}$	\bar{y}
$\hat{\theta}_{\text{CH}}$	1	-0.445	0.097	0.171	0.490	-0.290
$CH(\hat{\theta}_{\text{CH}})$	-0.445	1	-0.240	-0.170	-0.630	0.786
$\bar{\theta}_n^*$	0.097	-0.240	1	0.787	0.321	-0.025
$\hat{\theta}_n = \bar{\theta}_n$	0.171	-0.170	0.787	1	0.304	0.020
$\hat{\theta}_{\text{M}}$	0.490	-0.630	0.321	0.304	1	-0.386
\bar{y}	-0.290	0.786	-0.025	0.020	-0.386	1

Tabelle 4.8: Empirische Korrelationsmatrix zu den Bruchpunktschätzwerten bei unbekannten Risikoniveaus, zum Maximum der Calinski-Harabasz-Statistik und zur Ausfallquote im Zeitablauf

aus dem Rahmen und erscheint daher auch im Sinne der Zeitstabilität weniger gut geeignet. Ergänzend sind in Tabelle 4.8 die empirischen Korrelationen zwischen den verschiedenen Schätzwerten (ohne die Konstante), dem Maximum der CH-Statistik $CH(\hat{\theta}_{\text{CH}})$ und der quartalsweisen Ausfallquote zusammengetragen. Besonders bemerkenswert sind dabei aus Sicht des Autors folgende Punkte:

- Die Dempfle-Stute- und die Plug-In-Schätzwerte für θ sind stark positiv korreliert. Das war zu erwarten, da die beiden Ansätze über die Plug-In-Schätzer für die Risikoniveaus verknüpft sind.
- Die Dempfle-Stute- und die Plug-In-Schätzwerte für θ sind praktisch unkorreliert mit der quartalsweisen Ausfallquote. Konjunkturelle Einflüsse scheinen also keine Rolle zu spielen.
- Das Maximum der CH-Statistik ist stark positiv mit der quartalsweisen Ausfallquote korreliert. Für das Vergleichskriterium ist also eine starke Abhängigkeit von konjunkturellen Einflüssen zu attestieren.

Trotz der unter anderem in Bemerkung 4.1 angesprochenen Schwächen der CH-Statistik soll sie auch hier zum „Güte“-Vergleich der verschiedenen Methoden verwendet werden. Abbildung 4.25 zeigt im Zeitablauf die Werte der CH-Statistik, die sich ergeben, wenn die alternativen Bruchpunktschätzwerte eingesetzt werden. Tabelle 4.9 enthält die zugehörigen Lage- und Streuungsmaßzahlen. Dabei zeigt sich in jedem einzelnen Quartal die folgende Ordnung unter den Schätzwerten:

$$CH(\hat{\theta}_{\text{CH}}) > CH(500) > CH(\hat{\theta}_n) \geq CH(\bar{\theta}_n^*) > CH(\hat{\theta}_{\text{M}}).$$

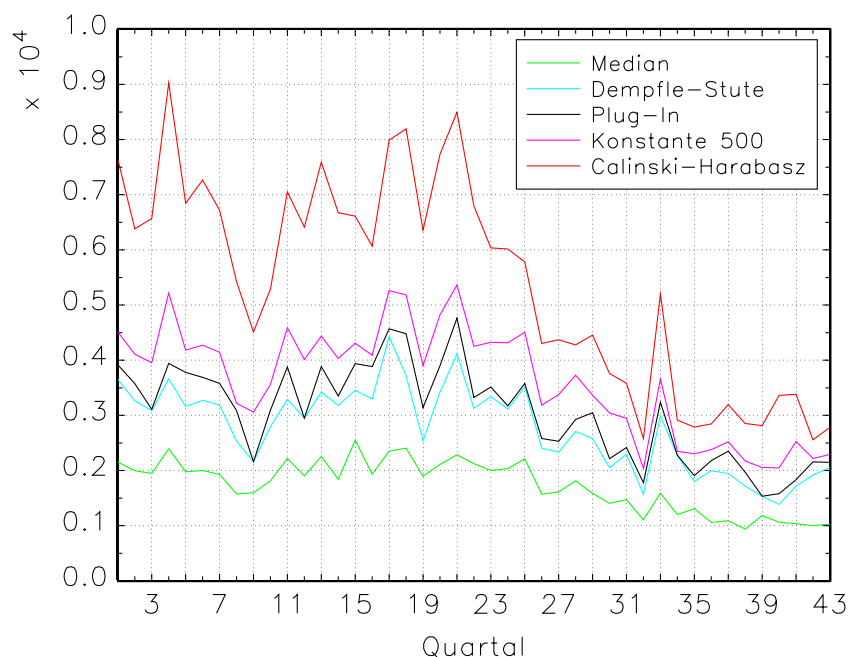


Abbildung 4.25: Werte der Calinski-Harabasz-Statistik für die Bruchpunktschätzwerte im Zeitablauf, unbekannte Risikoniveaus

	Calinski-Harabasz $CH(\hat{\theta}_{CH})$	Konstante $CH(500)$	Plug-In $CH(\bar{\theta}_n^*)$	Dempfle-Stute $CH(\hat{\theta}_n)$	Median $CH(\hat{\theta}_M)$
arithm. Mittel	5382.8	3623.3	3043.9	2766.0	1733.4
Median	5780.4	3902.8	3107.3	2949.1	1840.1
Minimum	2558.2	2035.9	1532.7	1383.7	938.1
in Quartal:	42	32	39	40	38
Maximum	9026.0	5360.3	4762.7	4425.5	2547.6
in Quartal:	4	21	21	17	15
Spannweite	6467.8	3324.5	3230.0	3041.8	1609.5
Standardabw.	1886.6	980.3	842.5	740.9	456.0

Tabelle 4.9: Lage- und Streuungsmaßzahlen zu den Werten der Calinski-Harabasz-Statistik, wenn die Klassen mit Hilfe der Bruchpunktschätzwerte bei unbekannten Risikoniveaus eingeteilt werden

	Calinski-Harabasz $F_n(\hat{\theta}_{CH})$	Konstante $F_n(500)$	Plug-In $F_n(\bar{\theta}_n^*)$	Dempfle-Stute $F_n(\hat{\theta}_n)$	Median $F_n(\hat{\theta}_M)$
arithm. Mittel	0.024	0.158	0.257	0.305	0.504
Median	0.020	0.158	0.245	0.303	0.502
Minimum	0.005	0.144	0.184	0.234	0.500
in Quartal:	8	43	42	42	40
Maximum	0.058	0.171	0.377	0.377	0.516
in Quartal:	42	7	39	39	41
Spannweite	0.053	0.027	0.193	0.143	0.016
Standardabw.	0.014	0.007	0.045	0.039	0.004

Tabelle 4.10: Lage- und Streuungsmaßzahlen zum Anteil der Ratingklasse 1 an allen Beobachtungen, wenn die Klassen mit Hilfe der Bruchpunktschätzwerte bei unbekannten Risikoniveaus eingeteilt werden

Würde die CH-Statistik als alleiniges Gütekriterium verwendet, dann müsste die Wahl logischerweise auf die Methode fallen, die die CH-Statistik maximiert. Des Weiteren würde in diesem Sinne der Plug-In-Schätzer besser als der Dempfle-Stute-Schätzer abschneiden. Dass die Verwendung der CH-Statistik als Gütekriterium jedoch nicht sinnvoll ist, zeigt sich daran, dass die naive Trennstelle 500 durchweg bessere Resultate liefert als die durchdachten und modellunterlegten Bruchpunktschätzwerte $\hat{\theta}_n$ und $\bar{\theta}_n^*$.

Die CH-Statistik disqualifiziert sich darüber hinaus auch im Hinblick auf die Besetzung der resultierenden Ratingklassen. Analog zu Abschnitt 4.2.1 betrachten wir wieder den Anteil der Beobachtungen in Ratingklasse 1 an der Gesamtzahl aller Beobachtungen. Die entsprechenden Lage- und Streuungsmaßzahlen sind in Tabelle 4.10 zusammengefasst. Dabei zeigt sich, dass bei der CH-Methodik im Mittel nur 2.4 %, im Extremfall sogar nur 0.5 % der Beobachtungen in Ratingklasse 1 liegen. Damit scheint nicht sichergestellt, dass der auf der Ratingklasseneinteilung beruhenden Schätzung der Ausfallwahrscheinlichkeiten eine ausreichende Datengrundlage zur Verfügung steht. Bei den anderen Methoden ist die Verteilung der Beobachtungen auf die beiden Ratingklassen aber hinreichend ausgewogen.

4.2.3 Konfidenzintervalle und Konfidenzregionen

Die Ziele dieses Abschnitts sind

- (a) die quartalsweise Bestimmung von asymptotischen Konfidenzintervallen $I_n(\gamma)$ für den unbekannten Bruchpunkt θ unter der Annahme bekannter Risikoniveaus gemäß (3.37), S. 64 und
- (b) die quartalsweise Bestimmung von simultanen asymptotischen Konfidenzregionen $C_n(\gamma)$ für die unbekannten Parameter θ , a und b gemäß (3.59), S. 79.

Bei einem Konfidenzniveau von $1 - \gamma$ benötigen wir

für (a) reelle Zahlen c_1 und d_1 , die die Ungleichung

$$P(\sigma < c_1) - P(\tau \leq -d_1) \geq 1 - \gamma \quad (4.9)$$

erfüllen und

für (b) reelle Zahlen c_1, c_2, c_3, d_1, d_2 und d_3 , die die Ungleichung

$$(P(\sigma < c_1) - P(\tau \leq -d_1))(\Phi(c_2) - \Phi(-d_2))(\Phi(c_3) - \Phi(-d_3)) \geq 1 - \gamma \quad (4.10)$$

erfüllen.

Dabei sind τ und σ die kleinste bzw. größte Maximalstelle des zusammengesetzten Poisson-Prozesses L aus (3.34) und Φ die Verteilungsfunktion der Standardnormalverteilung.

Wird im zweiten Fall das Konfidenzniveau $1 - \gamma$ gleichmäßig auf die drei Faktoren aufgeteilt, so resultieren aus (4.10) die folgenden drei Ungleichungen:

$$P(\sigma < c_1) - P(\tau \leq -d_1) \geq \sqrt[3]{1 - \gamma}, \quad (4.11)$$

$$\Phi(c_2) - \Phi(-d_2) \geq \sqrt[3]{1 - \gamma}, \quad (4.12)$$

$$\Phi(c_3) - \Phi(-d_3) \geq \sqrt[3]{1 - \gamma}. \quad (4.13)$$

Die Ungleichungen (4.12) und (4.13) sind erfüllt und in ihnen gilt sogar das Gleichheitszeichen, wenn

$$c_2 = c_3 = d_2 = d_3 = \Phi^{-1} \left(\frac{1 + \sqrt[3]{1 - \gamma}}{2} \right)$$

gesetzt wird,³¹ denn für $i = 2, 3$ gilt

$$\begin{aligned}
 \Phi(c_i) - \Phi(-d_i) &= \Phi\left(\Phi^{-1}\left(\frac{1+\sqrt[3]{1-\gamma}}{2}\right)\right) - \Phi\left(-\Phi^{-1}\left(\frac{1+\sqrt[3]{1-\gamma}}{2}\right)\right) \\
 &= \frac{1+\sqrt[3]{1-\gamma}}{2} - \left(1 - \Phi\left(\Phi^{-1}\left(\frac{1+\sqrt[3]{1-\gamma}}{2}\right)\right)\right) \\
 &= \frac{1+\sqrt[3]{1-\gamma}}{2} - 1 + \frac{1+\sqrt[3]{1-\gamma}}{2} \\
 &= \sqrt[3]{1-\gamma}.
 \end{aligned}$$

Beispielsweise ist

$$\Phi^{-1}\left(\frac{1+\sqrt[3]{1-\gamma}}{2}\right) \approx \begin{cases} 2.3877, & \text{falls } 1-\gamma = 0.95, \\ 2.9342, & \text{falls } 1-\gamma = 0.99. \end{cases}$$

Damit können die zwei zu den Risikoniveaus a und b gehörigen Teile der Konfidenzregion $C_n(\gamma)$ berechnet werden.

Für die Bestimmung der Konfidenzintervalle $I_n(\gamma)$ gemäß (3.37) bzw. der Konfidenzregionen $C_n(\gamma)$ gemäß (3.59) fehlen jetzt noch die Zahlen c_1 und d_1 in (4.9) bzw. in (4.11). Wir stellen fest, dass diese beiden Ungleichungen strukturell von der gleichen Form sind:

$$P(\sigma < c_1) - P(\tau \leq -d_1) \geq \nu.$$

Zur Wahl der Zahlen c_1 und d_1 wäre es wünschenswert, die Verteilungen von τ und σ zu kennen. Die Kenntnis der Verteilungen impliziert die Kenntnis aller Quantile. Würde dann für c_1 das $\frac{1+\nu}{2}$ -Quantil der Verteilung von σ und für $-d_1$ das $\frac{1-\nu}{2}$ -Quantil der Verteilung von τ gewählt,³² d. h.

$$c_1 = F_\sigma^{\leftarrow}\left(\frac{1+\nu}{2}\right) \quad \text{und} \quad -d_1 = F_\tau^{\leftarrow}\left(\frac{1-\nu}{2}\right),$$

und wären F_σ in c_1 und F_τ in $-d_1$ stetig, dann wären die Ungleichungen (4.9) und (4.11) erfüllt:

$$\begin{aligned}
 P(\sigma < c_1) - P(\tau \leq -d_1) &= F_\sigma(c_1) - F_\tau(-d_1) \\
 &= F_\sigma\left(F_\sigma^{\leftarrow}\left(\frac{1+\nu}{2}\right)\right) - F_\tau\left(F_\tau^{\leftarrow}\left(\frac{1-\nu}{2}\right)\right) \\
 &= \frac{1+\nu}{2} - \frac{1-\nu}{2} \\
 &= \nu.
 \end{aligned}$$

³¹Vgl. auch Ferger (2009, S. 81). Dabei bezeichnet Φ^{-1} die Quantilfunktion von Φ . Wegen der Stetigkeit und strengen Monotonie von Φ ist Φ^{-1} sogar die Umkehrfunktion (Inverse) von Φ .

³²Vgl. auch Ferger (2009, S. 81).

	$\frac{1 \pm \nu}{2}$	$1 - \gamma = 95 \%$	$1 - \gamma = 99 \%$
(a) $\nu = 1 - \gamma$	$\frac{1+(1-\gamma)}{2} = 1 - \frac{\gamma}{2}$	0.975	0.025
	$\frac{1-(1-\gamma)}{2} = \frac{\gamma}{2}$	0.995	0.005
(b) $\nu = \sqrt[3]{1-\gamma}$	$\frac{1+\sqrt[3]{1-\gamma}}{2}$	≈ 0.9915	≈ 0.9983
	$\frac{1-\sqrt[3]{1-\gamma}}{2}$	≈ 0.0085	≈ 0.0017

Tabelle 4.11: Bei der Quantilbestimmung interessierende Wahrscheinlichkeiten

Für die Konfidenzniveaus $1 - \gamma = 95 \%$ und $1 - \gamma = 99 \%$ sind dann die Quantile zu den in Tabelle 4.11 angegebenen Wahrscheinlichkeiten von Interesse.

Da die Verteilungen und daher auch die Quantile von τ und σ unbekannt sind, müssen letztere mit Hilfe einer Simulation geschätzt werden. Dabei werden die expliziten Darstellungen der Zufallsvariablen τ und σ aus Bemerkung 3.29, S. 65 verwendet. Als Eingangsparameter der Simulation dienen³³

für (a) die bekannten Risikoniveaus a und b sowie die geschätzte Intensität der Poisson-Prozesse $\lambda_n^* = \hat{f}_h(\theta_n^*)$ und

für (b) die Schätzwerte für die Risikoniveaus \bar{a}_n^* und \bar{b}_n^* sowie die geschätzte Intensität der Poisson-Prozesse $\bar{\lambda}_n = \hat{f}_h(\bar{\theta}_n)$.

Hierbei bezeichnet \hat{f}_h den in Gleichung (4.1), S. 88 vorgestellten Kerndichteschätzer. Wir nehmen der Einfachheit halber an, dass dieser auf einer Umgebung von θ fast sicher gleichmäßig gegen die wahre Dichte f konvergiert.³⁴ Eine Zusammenfassung der Lage- und Streuungsmaßzahlen der quartalsweisen Intensitätsschätzwerte ist in Tabelle 4.12 zu finden. Zu Vergleichszwecken sind in ihr auch Angaben zu den zu $\bar{\lambda}_n$ alternativen Schätzwerten $\bar{\lambda}_n^* = \hat{f}_h(\bar{\theta}_n^*)$ enthalten.

Für die Simulation müssen neben den Eingangsparametern auch die Anzahl der Monte-Carlo-Replikationen \mathcal{N} und die Indizes k_{\max} und m_{\max} , bis zu denen die Partialsommen V_k und W_m aus (3.38) ausgewertet werden, festgelegt werden. Dazu wurden Voruntersuchungen mit den Eingangsparametern

(a) $a = 0.01$, $b = 0.001$ und $\lambda = 0.0018$ bzw.

(b) $a = 0.0080$, $b = 0.0018$ und $\lambda = 0.0020$

durchgeführt. Die gewählten Parameter sind die als bekannt vorausgesetzten Risikoniveaus bzw. entsprechen etwa den zeitlichen Mitteln der zugehörigen geschätzten Größen

³³Vgl. Abschnitt 3.3, S. 65 bzw. Abschnitt 3.4, S. 77 oder Ferger & Klotsche (2009, S. 118).

³⁴Zur fast sicher gleichmäßigen Konvergenz des Kerndichteschätzers gegen die unbekannte Dichte f auf ganz \mathbb{R} vgl. z. B. Wied & Weißbach (2012), Silverman (1994, S. 71-72), Silverman (1978) oder Schuster (1969). Die Kriterienkataloge unterscheiden sich bei den Voraussetzungen an die Kernfunktion und an die Bandbreite. Die gleichmäßige Stetigkeit von f ist bei allen eine notwendige Annahme.

	$\lambda_n^* = \hat{f}_h(\theta_n^*)$	$\bar{\lambda}_n = \hat{f}_h(\bar{\theta}_n)$	$\bar{\lambda}_n^* = \hat{f}_h(\bar{\theta}_n^*)$
arithm. Mittel	0.00178	0.00201	0.00168
Median	0.00167	0.00195	0.00158
Minimum	0.00119	0.00155	0.00118
in Quartal:	8	34	16
Maximum	0.00290	0.00267	0.00267
in Quartal:	13	39	39
Spannweite	0.00171	0.00112	0.00149
Standardabw.	0.00038	0.00029	0.00032

Tabelle 4.12: Lage- und Streuungsmaßzahlen zu den geschätzten Intensitäten des zusammengesetzten Poisson-Prozesses

(siehe Tabelle 4.6 und Tabelle 4.12). Bei der Voruntersuchung zeigte sich, dass die Partialsummen V_k und W_m nur sehr langsam gegen $-\infty$ divergieren.³⁵ Abbildung 4.26 gibt davon einen Eindruck. Bei $m_{\max} = k_{\max} = 10\,000$ und einer Million Monte-Carlo-Replikationen waren die größten bei der Maximalstellensuche relevanten Indizes

(a) $\bar{k} = 3710$ und $\bar{m} = 3977$ bzw.

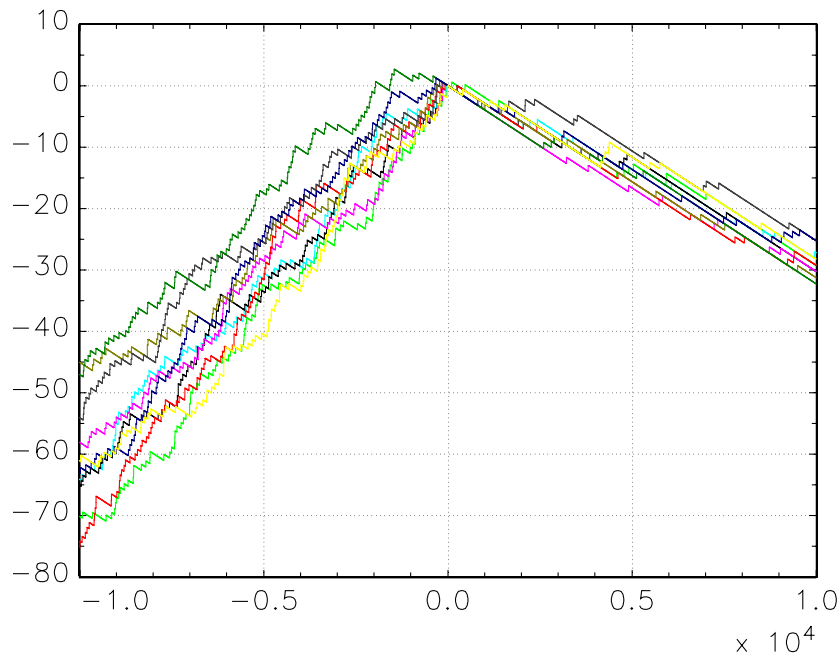
(b) $\bar{k} = 7393$ und $\bar{m} = 8066$.

Daher werden die Partialsummenprozesse V_k und W_m bei der quartalsweisen Ermittlung der Konfidenzintervalle bzw. -regionen bis zum Index 10 000 ausgewertet.

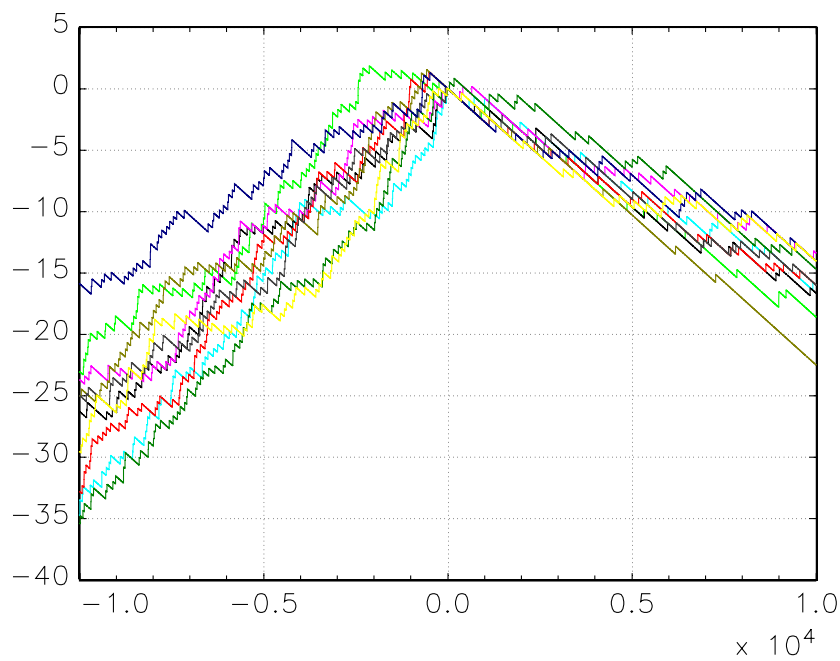
Der benötigte zusammengesetzte Poisson-Prozess unterscheidet sich von dem Partialsummenprozess in Abbildung 4.26 darin, dass er nicht in allen ganzen Zahlen $z \in \mathbb{Z} \setminus \{0\}$ springt, sondern die Sprungstellen „gestreckt“ sind. Die Zwischenereigniszeiten sind dabei mit Intensität λ exponentialverteilt. Bei der Untersuchung der kleinsten und größten Maximalstelle des zusammengesetzten Poisson-Prozesses, bezeichnet mit τ bzw. σ , ergaben sich die in Tabelle 4.13 angegebenen Lage- und Streuungsmaßzahlen. Es zeigt sich, dass die Kennzahlen bereits bei 20 000 Monte-Carlo-Replikationen in derselben Größenordnung liegen wie bei einer Million Simulationsschritten. Das gilt auch für die interessierenden Quantile in den Verteilungsenden. Da die Zahlen c_1 und d_1 bei der Berechnung des Konfidenzbereichs für θ zudem durch den Stichprobenumfang n , der im Datensatz jeweils größer als 1.1 Millionen ist,³⁶ geteilt werden, spielen die Unterschiede keine bedeutende Rolle. Insofern genügt bei der quartalsweisen Simulation die kleinere Replikationsanzahl. An dieser Stelle ist die Voruntersuchung beendet und wir ermitteln die konkreten Konfidenzintervalle bzw. -regionen.

³⁵Für diese und die folgenden Simulationen wurde das Software-Paket GAUSS, Version 12 eingesetzt.

³⁶Vgl. Abbildung 4.1 auf S. 84.



(a) $a = 0.01$, $b = 0.001$, $\lambda = 0.0018$



(b) $a = 0.0080$, $b = 0.0018$, $\lambda = 0.0020$

Abbildung 4.26: Je zehn Realisationen des aus den Partialsummen W_{-m} (links) und V_k (rechts) zusammengesetzten Prozesses, $m_{\max} = k_{\max} = 10\,000$

\mathcal{N}			20 000	100 000	1 000 000
(a) $a = 0.01$ $b = 0.001$ $\lambda = 0.0018$	τ	arithm. Mittel	−54 070	−53 863	−53 398
		Median	−36 487	−36 139	−35 667
		Standardabw.	142 732	142 892	143 115
		0.025-Quantil	−392 211	−392 817	−387 340
		0.005-Quantil	−639 483	−641 645	−644 551
	σ	arithm. Mittel	−53 516	−53 301	−52 840
		Median	−35 893	−35 571	−35 103
		Standardabw.	142 730	142 892	143 116
		0.975-Quantil	226 202	227 482	228 076
		0.995-Quantil	479 824	462 375	467 245
(b) $a = 0.0080$ $b = 0.0018$ $\lambda = 0.0020$	τ	arithm. Mittel	−68 841	−68 805	−67 979
		Median	−41 920	−42 526	−42 355
		Standardabw.	282 026	279 002	279 138
		0.0085-Quantil	−1 035 853	−1 036 445	−1 032 545
		0.0017-Quantil	−1 565 553	−1 538 929	−1 561 119
	σ	arithm. Mittel	−68 339	−68 305	−67 478
		Median	−41 306	−42 051	−41 848
		Standardabw.	282 024	279 002	279 139
		0.9915-Quantil	822 721	800 950	809 707
		0.9983-Quantil	1 353 225	1 312 443	1 312 912

Tabelle 4.13: Lage- und Streuungsmaßzahlen der durch Simulation ermittelten kleinsten und größten Maximalstelle τ bzw. σ des zusammengesetzten Poisson-Prozesses (Werte gerundet), verschiedene Eingangsparameter, unterschiedliche Anzahl \mathcal{N} von Monte-Carlo-Replikationen

Quartal	Konfidenzniveau $1 - \gamma$	Konfidenzintervall $I_n(\gamma)$
1	95 %]579.816, 580.314[
	99 %]579.635, 580.510[
21	95 %]607.849, 608.249[
	99 %]607.703, 608.401[
43	95 %]560.782, 561.372[
	99 %]560.543, 561.618[

Tabelle 4.14: Konfidenzintervalle für θ bei bekannten Risikoniveaus $a = 0.01$ und $b = 0.001$ für drei ausgewählte Quartale und zwei verschiedene Konfidenzniveaus

Konfidenzintervalle für θ gemäß (3.37), d. h.

$$I_n(\gamma) = \left] \theta_n^* - \frac{c_1}{n}, \theta_n^* + \frac{d_1}{n} \right[,$$

werden für die Konfidenzniveaus $1 - \gamma = 95\%$ und $1 - \gamma = 99\%$ berechnet. Für die Simulation der Zahlen c_1 und d_1 werden dabei die als bekannt vorausgesetzten Risikoniveaus $a = 0.01$ und $b = 0.001$, die im jeweiligen Quartal geschätzten Intensitäten $\lambda_n^* = \hat{f}_h(\theta_n^*)$, ferner $m_{\max} = k_{\max} = 10\,000$ sowie eine Replikationsanzahl von $\mathcal{N} = 20\,000$ gewählt. Für die Quartale 1, 21 und 43 sind die konkreten Konfidenzintervalle für θ in Tabelle 4.14 angegeben. Eine grafische Darstellung aller ermittelten Konfidenzintervalle im Zeitablauf hat wenig Sinn, da die Breite der Konfidenzintervalle im Vergleich zur Schwankungsbreite der Bruchpunktschätzwerte³⁷ verschwindend gering ist. Die Abweichung der Konfidenzintervallgrenzen vom zugehörigen Bruchpunktschätzwert beträgt selbst bei einem Konfidenzniveau von 99 % im extremsten Fall noch nicht einmal einen Scorepunkt. Vergleiche dazu Abbildung 4.27, in der die Abweichungen

$$\Delta^o := \left(\theta_n^* + \frac{d_1}{n} \right) - \theta_n^* = \frac{d_1}{n} \quad \text{und} \quad \Delta^u := \left(\theta_n^* - \frac{c_1}{n} \right) - \theta_n^* = -\frac{c_1}{n}$$

der oberen bzw. unteren Intervallgrenze vom ML-Schätzwert für alle Quartale und zwei Konfidenzniveaus visualisiert sind. In dieser Grafik sieht man auch, dass die Abweichungen nicht symmetrisch zu null sind, d. h. dass die Konfidenzintervalle nicht symmetrisch um den Schätzwert liegen.

Simultane Konfidenzregionen für (θ, a, b) gemäß (3.59), d. h.

$$C_n(\gamma) = \left] \bar{\theta}_n^* - \frac{c_1}{n}, \bar{\theta}_n^* + \frac{d_1}{n} \right[\times \left] \bar{a}_n^* - \frac{\bar{\sigma}_U}{\sqrt{n}} c_2, \bar{a}_n^* + \frac{\bar{\sigma}_U}{\sqrt{n}} d_2 \right[\times \left] \bar{b}_n^* - \frac{\bar{\sigma}_V}{\sqrt{n}} c_3, \bar{b}_n^* + \frac{\bar{\sigma}_V}{\sqrt{n}} d_3 \right[,$$

³⁷Vgl. Abbildung 4.18 auf S. 106 und Tabelle 4.2 auf S. 106.

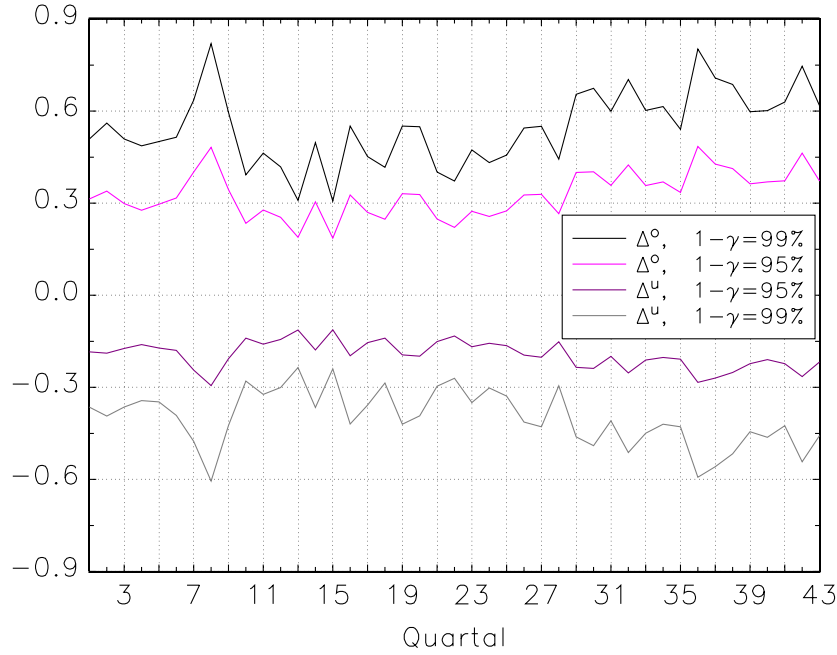


Abbildung 4.27: Abweichungen der Konfidenzintervallgrenzen von den jeweiligen Bruchpunktschätzwerten im Zeitablauf, bekannte Risikoniveaus

werden ebenfalls für die Konfidenzniveaus $1 - \gamma = 95 \%$ und $1 - \gamma = 99 \%$ berechnet. Dabei werden für die Simulation der Zahlen c_1 und d_1 die im jeweiligen Quartal geschätzten Größen \bar{a}_n^* , \bar{b}_n^* und $\bar{\lambda}_n = \hat{f}_h(\bar{\theta}_n)$, ferner $m_{\max} = k_{\max} = 10\,000$ sowie eine Replikationsanzahl von $\mathcal{N} = 20\,000$ verwendet. Die konkreten Konfidenzregionen sind beispielhaft für die Quartale 1, 21 und 43 in Tabelle 4.15 nachzulesen. Da die Seitenlängen der ermittelten Quader im Vergleich zur Schwankungsbreite der Bruchpunktschätzwerte bzw. der Teilausfallquoten verschwindend gering sind,³⁸ hat auch hier nur eine Darstellung der quartalsweisen Abweichungen

$$\begin{aligned}
 \Delta_1^o &:= \left(\bar{\theta}_n^* + \frac{d_1}{n} \right) - \bar{\theta}_n^* = \frac{d_1}{n}, & \Delta_1^u &:= \left(\bar{\theta}_n^* - \frac{c_1}{n} \right) - \bar{\theta}_n^* = -\frac{c_1}{n}, \\
 \Delta_2^o &:= \left(\bar{a}_n^* + \frac{\bar{\sigma}_U}{\sqrt{n}} d_2 \right) - \bar{a}_n^* = \frac{\bar{\sigma}_U}{\sqrt{n}} d_2, & \Delta_2^u &:= \left(\bar{a}_n^* - \frac{\bar{\sigma}_U}{\sqrt{n}} c_2 \right) - \bar{a}_n^* = -\frac{\bar{\sigma}_U}{\sqrt{n}} c_2, \\
 \Delta_3^o &:= \left(\bar{b}_n^* + \frac{\bar{\sigma}_V}{\sqrt{n}} d_3 \right) - \bar{b}_n^* = \frac{\bar{\sigma}_V}{\sqrt{n}} d_3, & \Delta_3^u &:= \left(\bar{b}_n^* - \frac{\bar{\sigma}_V}{\sqrt{n}} c_3 \right) - \bar{b}_n^* = -\frac{\bar{\sigma}_V}{\sqrt{n}} c_3
 \end{aligned}$$

einen Sinn (vgl. Abbildung 4.28). Hierbei wird z. B. deutlich, dass die Breite des zum Bruchpunkt θ gehörigen Teils der simultanen Konfidenzregion zum Konfidenzniveau 99% in jedem Quartal kleiner als vier Score-Einheiten ist. Bei einem Konfidenzniveau von 95% beträgt die entsprechende Breite maximal 2.51 Einheiten.

³⁸Vgl. Abbildung 4.24 auf S. 116 und Tabelle 4.7 auf S. 116 bzw. Abbildung 4.22 auf S. 114 und Tabelle 4.6 auf S. 114.

Quartal	Konfidenzniveau	simultane Konfidenzregion
	$1 - \gamma$	$C_n(\gamma)$
1	95 %	$]557.532, 558.675[\times]0.0089563, 0.0090957[\times]0.0015638, 0.0015734[$
	99 %	$]557.196, 559.047[\times]0.0089403, 0.0091116[\times]0.0015627, 0.0015745[$
21	95 %	$]536.496, 537.693[\times]0.0098597, 0.0100149[\times]0.0019830, 0.0019940[$
	99 %	$]536.117, 537.994[\times]0.0098420, 0.0100327[\times]0.0019818, 0.0019952[$
43	95 %	$]560.028, 562.242[\times]0.0069324, 0.0070604[\times]0.0016375, 0.0016472[$
	99 %	$]559.350, 562.900[\times]0.0069178, 0.0070750[\times]0.0016364, 0.0016483[$

Tabelle 4.15: Konfidenzregionen für (θ, a, b) bei unbekannten Risikoniveaus für drei ausgewählte Quartale und zwei verschiedene Konfidenzniveaus

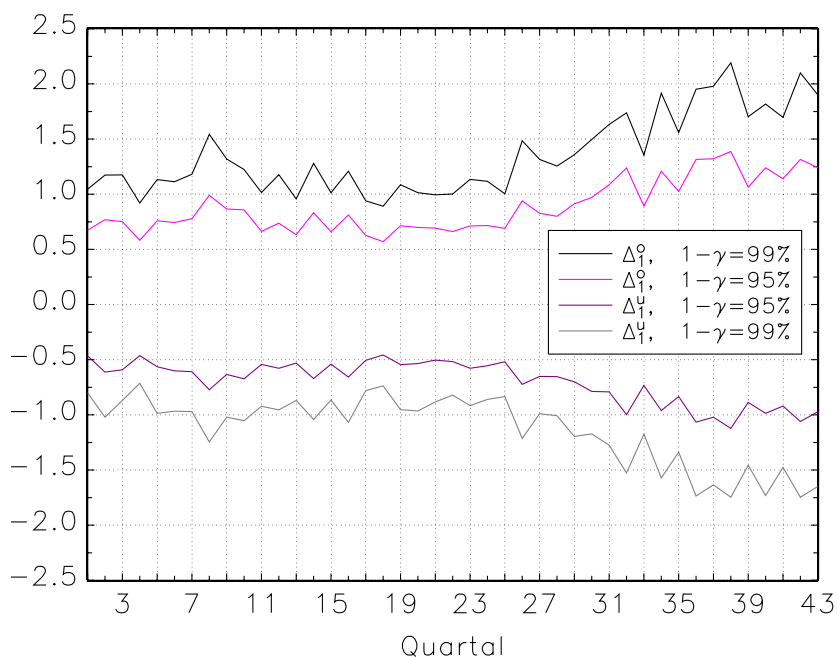
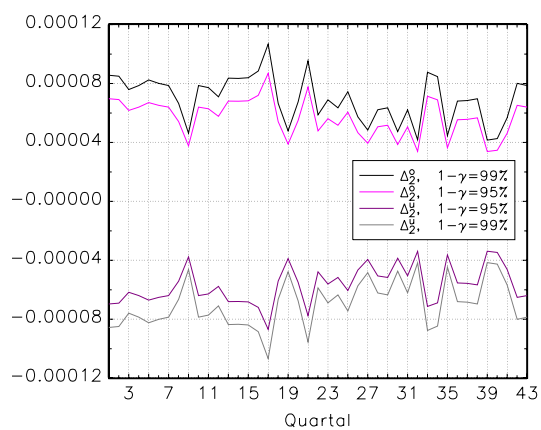
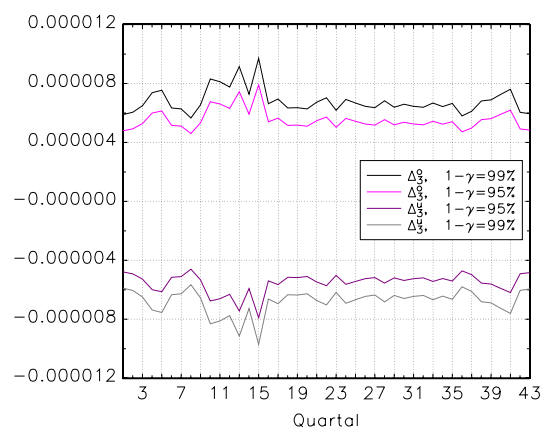
(a) Bruchpunkt θ (b) Risikoniveau a (c) Risikoniveau b

Abbildung 4.28: Abweichungen der Intervallgrenzen der simultanen Konfidenzregionen von den jeweiligen Schätzwerten im Zeitablauf, unbekannte Risikoniveaus

4.3 Bruchpunktschätzung im Pool aller Quartale

In Abschnitt 4.2 wurden die Schätzwerte für die gemeinsame Grenze der Ratingklassen quartalsweise ermittelt. Obwohl die modellunterlegten Bruchpunktschätzwerte im Vergleich zu den alternativen fortgeschrittenen Ansätzen recht zeitstabil waren, zeigten sich doch große Spannweiten. Die Schwankungsbreite betrug für die Maximum-Likelihood-Schätzwerte bei vorgegebenen Risikoniveaus 129 Scorepunkte, bei den Plug-In-Schätzwerten 110 und bei den Dempfle-Stute-Schätzwerten immerhin noch 69 Scorepunkte.³⁹ Daher werden im Folgenden Schätzwerte für die Trennstelle der Klassen ermittelt, welche auf den Daten aller 43 Quartale fußen. Die Verwendung des gesamten Datenpools genügt eher den Anforderungen der Basler Richtlinien. Diese erwarten, dass sich die Ratingklassenbildung auf einen größeren Zeithorizont stützt.⁴⁰ Durch das Einbeziehen aller Quartale erhalten wir einen Stichprobenumfang von $n = \sum_{t=1}^{43} n_t = 53\,629\,537$. Es ist allerdings zu beachten, dass gleiche Personen mehrmals im Gesamtdatensatz auftauchen und dabei auch verschiedene Scores besitzen können. Insofern bleibt die Unabhängigkeitsannahme weiterhin kritisch. Man kann jedoch davon ausgehen, dass durch die Längsschnittbetrachtung die Schätzwerte für die Risikoniveaus, welche insbesondere für den Plug-In-Schätzer benötigt werden, besser sind als in der reinen Querschnittsuntersuchung.⁴¹

Widmen wir uns zuerst dem Maximum-Likelihood-Schätzer θ_n^* . Dieser setzt die Kenntnis der Risikoniveaus a und b voraus.

- Hier wählen wir zum einen, genau wie in Abschnitt 4.2.1, die Werte $a = 0.01$ und $b = 0.001$. Es resultiert der Schätzwert

$$\theta_n^* = \operatorname{argmax}_{x \in \mathbb{R}} S_n^*(x) = 581.$$

Dieser entspricht etwa dem arithmetischen Mittel der quartalsweise ermittelten ML-Schätzwerte.⁴²

- Zum anderen verwenden wir zu Vergleichszwecken die Werte $a = 0.00797$ und $b = 0.00180$. Diese entsprechen etwa den zeitlichen Mitteln der Teilausfallquoten aus Abschnitt 4.2.2, Tabelle 4.6.⁴³ In diesem Fall resultiert der Schätzwert

$$\theta_n^* = \operatorname{argmax}_{x \in \mathbb{R}} S_n^*(x) = 563.$$

Die zugehörigen empirischen Prozesse S_n^* sind in Abbildung 4.29 visualisiert.

³⁹Vgl. Tabelle 4.2 bzw. Tabelle 4.7.

⁴⁰Vgl. Abschnitt 2.2.

⁴¹Vgl. Abschnitt 2.3.

⁴²Vgl. Tabelle 4.2.

⁴³Wird anstelle der zeitlichen Mittel jeweils der Median der Ausfallquoten verwendet, d. h. $a = 0.00828$ und $b = 0.00171$, resultiert derselbe Bruchpunktschätzwert.

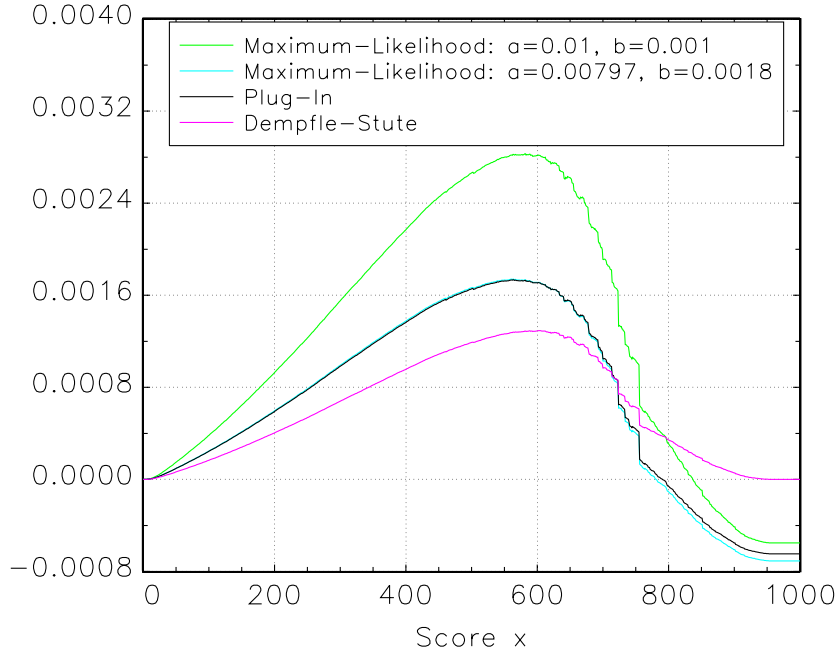


Abbildung 4.29: Funktionsverlauf der empirischen Prozesse S_n^* (Maximum-Likelihood), S_n (Dempfle-Stute) und \bar{S}_n^* (Plug-In) im Pool aller Quartale

Der Funktionsverlauf des empirischen Prozesses S_n , auf dem die Dempfle-Stute-Schätzer basieren, ist ebenfalls in dieser Abbildung zu sehen. Offensichtlich ist der Ausschlag der Funktion S_n in die positiven Zahlen größer als in die negativen, d. h. $\max_{x \in \mathbb{R}} S_n(x) > \max_{x \in \mathbb{R}} -S_n(x)$, so dass gemäß (3.22) der ein- und der zweiseitige Dempfle-Stute-Schätzwert identisch sind. Konkret ergibt sich als kleinste Maximalstelle des empirischen Prozesses und damit als Bruchpunktschätzwert

$$\bar{\theta}_n = \hat{\theta}_n = \operatorname{argmax}_{x \in \mathbb{R}} S_n(x) = 602.$$

Dies ist gerade der Median der quartalsbezogenen Dempfle-Stute-Schätzwerte.⁴⁴

Auf Basis des Dempfle-Stute-Schätzwerts 602 lassen sich für die unbekannten Risikoniveaus a und b die Plug-In-Schätzwerte

$$\bar{a}_n^* = a_n^*(602) \approx 0.0078501 \quad \text{und} \quad \bar{b}_n^* = b_n^*(602) \approx 0.0018030$$

ermitteln. Werden diese Zahlen als Eingangsparameter der ML-Schätzung verwendet, resultiert der Plug-In-Schätzwert für den Bruchpunkt θ . Dieser Schätzwert fußt auf dem empirischen Prozess \bar{S}_n^* , dessen Funktionsverlauf ebenfalls in Abbildung 4.29 dargestellt ist. Es gilt

$$\bar{\theta}_n^* = \operatorname{argmax}_{x \in \mathbb{R}} \bar{S}_n^*(x) = 563.$$

⁴⁴Vgl. Tabelle 4.7.

Auch hier ist der Pool-Schätzwert gleich dem Median der quartalsweise berechneten Plug-In-Schätzwerte.⁴⁵

Der Dempfle-Stute-Schätzwert und der Plug-In-Schätzwert liegen – mutmaßlich wegen nicht zutreffender Modellannahmen – weit auseinander. Ihre Differenz beträgt 39 Scorepunkte. Wären alle entsprechenden Annahmen erfüllt, würde man aufgrund der Konsistenz der Schätzer erwarten, dass die Schätzwerte bei einem dermaßen großen Stichprobenumfang viel näher beieinander liegen. Auch aus dieser Überlegung heraus wird an dieser Stelle auf die Angabe von konkreten Konfidenzintervallen oder -regionen verzichtet. Die Konfidenzbereiche würden der Schätzung aufgrund des gegenüber der Quartalsbetrachtung etwa um den Faktor 40 vergrößerten Stichprobenumfangs eine Genauigkeit zuschreiben, die nicht gegeben zu sein scheint.

Zum Abschluss dieses Abschnitts prüfen wir, ob die auf den Pool-Schätzwerten basierenden Ratingsysteme sinnvolle Schätzwerte für die ratingklassenspezifischen Ausfallwahrscheinlichkeiten generieren. Teilen wir dazu die Scores beispielsweise mit dem Dempfle-Stute-Schätzwert des Pools in zwei Klassen ein, besteht Ratingklasse 1 aus den Scores von 1 bis 602 und Ratingklasse 2 aus den Scores von 603 bis 1000. Nun werden quartalsweise die Ausfallwahrscheinlichkeiten a und b der beiden Ratingklassen geschätzt. Die Schätzung erfolgt mit den Teilausfallquoten⁴⁶

$$\hat{a}_t := a_{n_t,t}^*(602) = \frac{\sum_{i=1}^{n_t} Y_{it} \mathbb{1}_{\{X_{it} \leq 602\}}}{\sum_{i=1}^{n_t} \mathbb{1}_{\{X_{it} \leq 602\}}}$$

und

$$\hat{b}_t := b_{n_t,t}^*(602) = \frac{\sum_{i=1}^{n_t} Y_{it} \mathbb{1}_{\{X_{it} > 602\}}}{\sum_{i=1}^{n_t} \mathbb{1}_{\{X_{it} > 602\}}}.$$

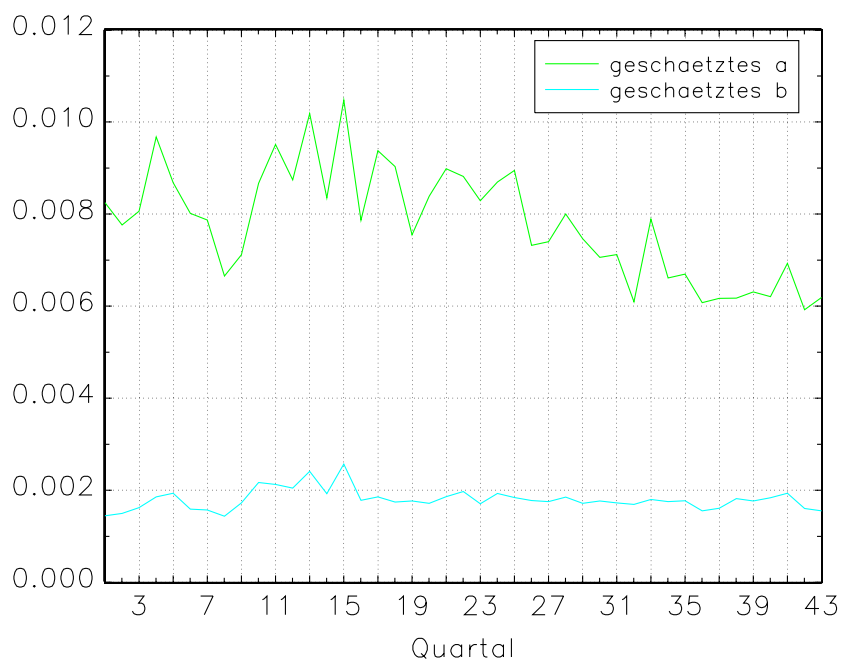
Die resultierenden Schätzwerte sind in Abbildung 4.30(a) dargestellt. Sie erfüllen offensichtlich die Ungleichungskette

$$\max_t \hat{a}_t \geq \boxed{\min_t \hat{a}_t > \max_t \hat{b}_t} \geq \min_t \hat{b}_t$$

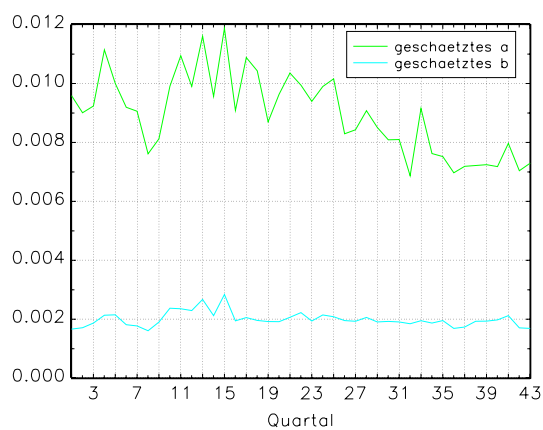
und damit die Forderung aus (2.22). Die periodenbezogenen Ausfallquoten bewegen sich jeweils in einem für die Ratingklasse typischen Bereich und die Bereiche überlappen sich nicht. Diese Bedingung ist auch erfüllt, wenn die Ratingklassen mit dem Pool-Plug-In-Schätzwert 563 oder mit dem Pool-Maximum-Likelihood-Schätzwert 581 gebildet werden. Vergleiche dazu Abbildung 4.30(b) bzw. 4.30(c). Weil nur zwei Ratingklassen gebildet wurden, wäre ein gegenteiliges Ergebnis auch sehr überraschend gewesen.

⁴⁵Vgl. Tabelle 4.7.

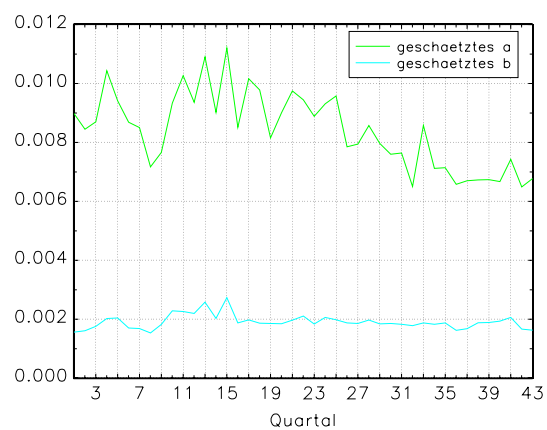
⁴⁶Vgl. die Gleichungen (3.46) und (3.47) auf S. 69.



(a) Dempfle-Stute, Trennstelle 602



(b) Plug-In, Trennstelle 563



(c) Maximum-Likelihood, Trennstelle 581

Abbildung 4.30: Zeitreihen der auf den Pool-Schätzwerten basierenden quartalsweisen Teilausfallquoten

Kapitel 5

Zusammenfassung und Ausblick

5.1 Zusammenfassung

Die vorliegende Arbeit befasste sich mit Einsatzmöglichkeiten der Bruchpunktschätzung bei der Bildung von Ratingklassen. Dabei wurde der Fall untersucht, dass genau zwei Ratingklassen gebildet werden sollen. In der Gesamtschau ist festzustellen, dass die Methoden der Bruchpunktschätzung geeignet sind, dieses Problem zu lösen. Sie unterscheiden sich von den Ansätzen aus Abschnitt 2.2 insbesondere dadurch, dass sie mit einem parametrischen Modell unterlegt sind (Abschnitt 3.1). Dadurch wird es möglich, Güteeigenschaften der Schätzer zu untersuchen.

Wird speziell der Fall bekannter Risikoniveaus betrachtet, d. h. sind für die Ratingklassen Zielausfallwahrscheinlichkeiten vorgegeben, kann der Maximum-Likelihood-Schätzer (ML-Schätzer) aus Abschnitt 3.3 verwendet werden. Wenn die entsprechenden Bedingungen erfüllt sind, ist dieser Schätzer stark konsistent (Satz 3.22) oder sogar n -konsistent (Satz 3.23). Unter etwas stärkeren Annahmen lässt sich ein asymptotisches Konfidenzintervall für den Bruchpunkt zwischen den Ratingklassen konstruieren (Folgerung 3.28).

Empirisch werden die vorgestellten Schätzer anhand eines Datensatzes einer deutschen Auskunftseinstellung untersucht (Kapitel 4). Hierbei bewegen sich die ML-Schätzwerte innerhalb der Grenzen der praxiserprobten Ansätze von Fernandes (2005) und Anderson (2007). Im Besonderen sind die ML-Schätzwerte stets größer als die Fernandes-Trennstellen und stets kleiner als die Anderson-Trennstellen. Dabei ist es unerheblich, ob diese im Original verwendet werden oder gemäß Bemerkung 2.4 modifiziert sind. In der quartalsbezogenen Betrachtung (Abschnitt 4.2.1) zeigte sich, dass die ML-Schätzwerte im Zeitverlauf erheblich schwanken. Im Vergleich zu den alternativen Methoden sind sie dennoch zeitstabil, und zwar sowohl in Bezug auf die Größenordnung der Trennstelle als auch im Hinblick auf die Besetzung der Ratingklassen. Die konkreten Konfidenzintervalle sind sehr schmal (Abschnitt 4.2.3). In der quartalsübergreifenden Betrachtung

(Abschnitt 4.3) bestätigt sich die Größenordnung der Bruchpunktschätzwerte. Die Validierung des resultierenden Ratingmodells gemäß dem Verfahren aus Abschnitt 2.3 ist erfolgreich: Die vierteljährlichen ratingklassenspezifischen Ausfallquoten überlappen sich im Zeitablauf nicht.

Sind neben dem Bruchpunkt auch die Risikoniveaus unbekannt, d. h. keine Ausfallwahrscheinlichkeiten als Referenz vorgegeben, können die Dempfle-Stute-Schätzer aus Abschnitt 3.2 oder die Plug-In-Schätzer aus Abschnitt 3.4 dazu verwendet werden, die Trennstelle zwischen den Ratingklassen zu schätzen. Die Schätzer sind stark konsistent oder sogar n -konsistent, wenn die Voraussetzungen der betreffenden Sätze erfüllt sind (Sätze 3.12 u. 3.16 bzw. Sätze 3.36 u. 3.38). Auf Basis des Plug-In-Schätzers lässt sich ferner ein asymptotisches Konfidenzintervall für den Bruchpunkt zwischen den Ratingklassen konstruieren (Folgerung 3.42). In diesem Zusammenhang ergeben sich auch Plug-In-Schätzer für die Risikoniveaus. Diese Schätzer sind stark konsistent und asymptotisch normal, wenn die Prämissen von Satz 3.33 erfüllt sind. Zusammen genommen lässt sich eine asymptotische Konfidenzregion für den Bruchpunkt und die Risikoniveaus angeben (Folgerung 3.44).

In der empirischen Untersuchung (Kapitel 4) wurden die Dempfle-Stute- und die Plug-In-Methode mit zwei naiven Kriterien und einem Optimierungsansatz, nämlich dem Calinski-Harabasz-Kriterium (CH-Kriterium) aus Abschnitt 2.2.2, verglichen. In der quartalsbezogenen Betrachtung (Abschnitt 4.2.2) zeigten sich dabei unerwartet große Unterschiede zwischen den Dempfle-Stute- und den Plug-In-Schätzwerten. Diese sind mutmaßlich darauf zurückzuführen, dass die Daten bezüglich ihrer Regressionsfunktion stark vom Modell abweichen. Auch wenn die Bruchpunktschätzwerte im Zeitablauf erheblich schwanken, sind sie doch viel stabiler als die Ergebnisse des CH-Kriteriums. Letzteres liefert zudem schwach besetzte Ratingklassen. Folglich halten wir fest, dass die CH-Statistik, angewendet auf Ausfallindikatoren und -quoten, als Kriterium zur Ratingklasseneinteilung ungeeignet ist (vgl. auch Bemerkung 4.1). Ebenso ungeeignet erscheint die CH-Statistik zum Vergleich von verschiedenen Ratingoptionen.

Ein bemerkenswertes Ergebnis der Untersuchung war, dass die Bruchpunktschätzwerte und die quartalsweisen Ausfallquoten im Zeitablauf praktisch unkorreliert sind. Das lässt vermuten, dass die Bruchpunktschätzer unabhängig von konjunkturellen Einflüssen arbeiten. Die Plug-In-Schätzwerte für die Risikoniveaus lassen eine analoge Vermutung dagegen nicht aufkommen. Sie streuen im starken Gleichlauf mit der Ausfallquote. Die simultane Schätzung von Bruchpunkt und Risikoniveaus liefert sehr kleine konkrete Konfidenzregionen (Abschnitt 4.2.3). Die quartalsübergreifende Berechnung von Dempfle-Stute- und Plug-In-Schätzwerten (Abschnitt 4.3) bestätigt schließlich die Größenordnung der quartalsbezogenen Werte. Die Validierung der resultierenden Ra-

tingmodelle gemäß dem Verfahren aus Abschnitt 2.3 verläuft erwartungsgemäß erfolgreich.

Zusammenfassend erweist sich die Bruchpunktschätzung sowohl aus theoretischer Sicht als auch in praktischer Hinsicht als vorteilhaft. Ersteres begründet sich in der Einführung eines parametrischen Modells und der dadurch ermöglichten Untersuchung der Güte der Parameterschätzer. Zweiteres zeigt sich in den unter verschiedenen Aspekten respektablen Ergebnissen bei der Anwendung auf einen Datensatz.

5.2 Ausblick

Trotz des positiven Fazits lässt die vorliegende Arbeit etliche Fragen offen. An manchen Stellen scheinen Erweiterungen oder Verallgemeinerungen ganz natürlich. In diesem letzten Abschnitt sollen daher mögliche Stoßrichtungen zukünftiger Forschung skizziert werden. Sie betreffen:

1. den Ausbau der Untersuchungen im Zweiklassenfall,
2. die Verallgemeinerung auf den K -Klassen-Fall und
3. die Erweiterung des Modells für den Fall abhängiger Beobachtungen.

Das erste große Forschungsgebiet liegt im Ausbau der Untersuchungen im Zweiklassenfall. Auf der theoretischen Seite gibt es dabei folgende Erweiterungsmöglichkeiten.

- In der Simulationsstudie von Ferger & Klotzsche (2009) wird hauptsächlich untersucht, wie sich die Lage des Bruchpunkts auf die entsprechenden Schätzer auswirkt. Die verwendeten Werte für die Risikoniveaus spielen eine untergeordnete Rolle. Im vorliegenden Kontext sind die von Ferger & Klotzsche benutzten Parameterkonstellationen unpassend, weil ihre Differenz untereinander und zu den Extrempositionen 0 und 1 zu groß ist. Insofern könnte die Studie gerade um extreme Risikoniveaus erweitert werden, um zu überprüfen, ab welchem Stichprobenumfang mit ausreichender Güte geschätzt werden kann (vgl. auch Bemerkung 3.45, S. 80).
- Basierend auf den Aussagen zur asymptotischen Verteilung der Bruchpunkt- und Risikoniveauschätzer könnten asymptotische Testverfahren für die Modellparameter konstruiert werden. Solche Verfahren könnten z. B. bei der Untersuchung sektoraler Unterschiede in Kreditausfalldaten¹ eingesetzt werden.

¹Vgl. zu diesem Problem Lehmann (2013).

- Für die Bruchpunktschätzer wurden die asymptotischen Güteeigenschaften bereits diskutiert. Somit stellt sich die Frage, welche Güteeigenschaften die alternativen Schätzer für den Bruchpunkt im vorliegenden Modell besitzen. Lassen sich Voraussetzungen angeben, unter denen diese Schätzer z. B. stark konsistent sind?

Auf der empirischen Seite entstehen im Zweiklassenfall die folgenden Fragen und Ideen.

- Die quartalsbezogenen Plug-In-Schätzwerte für den Bruchpunkt basieren in der vorliegenden Arbeit auf den Dempfle-Stute-Schätzwerten und den daraus resultierenden quartalsweisen Teilausfallquoten. In diesem Zusammenhang ergeben sich nachstehende Fragen:
 - Wie verändern sich die Plug-In-Schätzwerte, wenn anstelle dieser einperiodigen Ausfallquoten die langfristigen Ausfallquoten verwendet werden?
 - Wie verändern sich die Plug-In-Schätzwerte, wenn anstelle der Dempfle-Stute-Schätzwerte die Plug-In-Schätzwerte der ersten Stufe verwendet werden? Wie verhalten sich die Plug-In-Schätzwerte, wenn diese Idee iterativ angewendet wird?
- Einige der in Abschnitt 2.2 vorgestellten Methoden zur Ratingklasseneinteilung konnten in der empirischen Untersuchung nicht berücksichtigt werden, weil sie die geschätzten individualspezifischen Ausfallwahrscheinlichkeiten erfordern. Diese liegen im betrachteten Datensatz nicht explizit vor. Um die Bruchpunktansätze auch mit diesen Methoden zu vergleichen, gibt es folgende Möglichkeiten:
 - Den Bonitätsscores des betrachteten Auskunft-Datensatzes liegen geschätzte individualspezifische Ausfallwahrscheinlichkeiten zugrunde. Eine Rücktransformation der Scores würde die benötigten Werte liefern.
 - Es könnte ein anderer Datensatz behandelt werden, der die nötigen Größen explizit beinhaltet.

Bei der Bruchpunktschätzung wird dann die geschätzte individualspezifische Ausfallwahrscheinlichkeit als erklärende Variable X verwendet. Davon ausgehend ergibt sich ein weiteres Kriterium zur Bestimmung der Trennstelle zwischen den Ratingklassen.

- Bei der Bruchpunktschätzung wurde in dieser Arbeit stets die dichotome Ausfallvariable als erklärte Variable Y genutzt. Dempfle & Stute (2002) lassen aber auch allgemeinere Regressionsmodelle zu. Wie im von Krink et al. (2008) vorgeschlagenen Ansatz (2.19), S. 26 könnte ebenfalls der Zinssatz,

der sich aus den geschätzten individuenspezifischen Ausfallwahrscheinlichkeiten ergibt, als Zielvariable herangezogen werden.

Bei den empirischen Untersuchungen bleibt aber stets das Problem bestehen, dass die Annahmen zur Gestalt der Regressionsfunktion κ und zur stochastischen Unabhängigkeit der Beobachtungen kritisch sind.

Das zweite Forschungsgebiet betrifft daher die Gestalt der Regressionsfunktion κ , genauer gesagt die Anzahl ihrer Stufen. Bisher wurde nur der Zweiklassenfall behandelt. Er basiert auf einer einstufigen Regressionsfunktion. Im Kontext der Ratingklasseneinteilung werden aber stets mehr als zwei Ratingklassen verwendet.² Eine erste Möglichkeit, um mehrere Klassen zu bilden, wäre es, die Bruchpunktschätzer aus dem Zweiklassenfall iterativ anzuwenden. Dieses Vorgehen führt aber nur zu lokal besten Lösungen. Außerdem stellt sich in jedem Schritt die Frage, welche bestehende Klasse als nächstes in zwei neue Klassen aufgeteilt wird. Diese Variante kann also nur eine provisorische Lösung sein. Sinnvoller wäre es, die Beobachtungen simultan in K Klassen einzuteilen. Dazu muss die Regressionsfunktion κ verallgemeinert werden. Man ersetze hierfür die Modellgleichung (3.2) durch

$$\kappa(x) = \begin{cases} p_1, & \text{falls } x \in]-\infty, \theta_1], \\ p_2, & \text{falls } x \in]\theta_1, \theta_2], \\ \vdots & \vdots \\ p_K, & \text{falls } x \in]\theta_{K-1}, +\infty[, \end{cases} \quad (5.1)$$

wobei für die Risikoniveaus

$$p_k \in]0, 1[, k = 1, \dots, K \quad \text{sowie} \quad p_k \neq p_{k+1}, k = 1, \dots, K - 1 \quad (5.2)$$

und für die Bruchpunkte

$$-\infty < \theta_1 < \theta_2 < \dots < \theta_{K-1} < +\infty$$

gelte. Die Bedingungen an die Risikoniveaus in (5.2) sind sehr allgemein gehalten. Es wird lediglich gefordert, dass benachbarte Klassen unterschiedliche Risikoniveaus besitzen. Im Anwendungsfall ist häufig mehr Information vorhanden. So erwartet man

²Der Basler Ausschuss für Bankenaufsicht (BCBS, 2006, S. 91, Paragraph 404) fordert für Firmen-, Staats- und Bankkredite mindestens sieben Ratingklassen für nicht ausgefallene Kreditnehmer. Pfingsten & Schröck (2000, S. 18) empfehlen ein mindestens acht- bis zehnstufiges Ratingsystem. Der Deutsche Sparkassen- und Giroverband setzt laut Fischer (2002) ein System mit 15 Klassen ein. Die Schufa verwendet laut Korczak & Wilken (2009) zwölf Ratingstufen für den Branchenscore für Sparkassen. Die Ratingagenturen Standard & Poor's, Moody's sowie Fitch benutzen ein mindestens neun Klassen umfassendes Ratingsystem (vgl. Bluhm et al., 2010, S. 6-7).

speziell bei Ratingsystemen, dass das Ausfallrisiko mit wachsender Bonität monoton abnimmt. Dementsprechend könnte dann statt (5.2) die folgende Bedingung gestellt werden:

$$1 > p_1 > p_2 > \dots > p_K > 0. \quad (5.3)$$

Problematisch bleibt weiterhin die Festlegung der Klassenzahl K , wenn diese nicht vorgegeben ist. Im hier betrachteten Anwendungsfall könnte man diese Zahl mit Hilfe der Validierungskriterien in Abschnitt 2.3 zumindest eingrenzen:

- (i) Es wird die Einhaltung von (2.22) gefordert, d. h. die Ausfallquoten der Ratingklassen dürfen sich im Zeitablauf nicht überlappen. Das kann insbesondere für die besten Klassen jedoch ein zu hartes Kriterium sein.
- (ii) Die langfristigen Ausfallquoten müssen über die Ratingklassen monoton sein, d. h. analog zu (5.3) soll $\bar{y}_1 > \bar{y}_2 > \dots > \bar{y}_K$ gelten. Ferner dürfen sich die kurzfristigen Ausfallquoten nicht zu häufig überschneiden.

Auch die resultierenden Klassenbesetzungen sollten berücksichtigt werden. Sind schließlich geeignete Bruchpunktschätzer für das Modell (3.1) mit κ gemäß (5.1) gefunden, könnten diese wiederum empirisch überprüft und mit den alternativen Ansätzen aus Abschnitt 2.2 verglichen werden.

Das dritte Forschungsgebiet ist vom vorliegenden Kontext des Kreditrisikos motiviert, in dem nicht von der stochastischen Unabhängigkeit der Beobachtungen ausgegangen werden kann. Somit gilt es, zusätzlich Abhängigkeit in das Modell zu integrieren. Für die Bonitätsvariablen könnte im Einklang mit der einschlägigen Literatur³ ein Einfaktormodell unterstellt werden:

$$X_i = \sqrt{\varrho} Z + \sqrt{1 - \varrho} U_i, \quad i = 1, \dots, n. \quad (5.4)$$

Dabei ist Z der auf alle Bonitätsvariablen X_i wirkende systematische Faktor. Die Zufallsvariablen U_i sind die individuenspezifischen oder idiosynkratischen Faktoren. Sie werden als stochastisch unabhängig und identisch verteilt angenommen und seien zudem stochastisch unabhängig von Z . Der Parameter $\varrho \in [0, 1[$ quantifiziert die Stärke der Abhängigkeit zwischen Bonitätsvariable und systematischem Faktor.⁴ Wird zusätz-

³Vgl. z. B. Henking et al. (2006, S. 163), Bluhm et al. (2010, S. 80) oder McNeil et al. (2005, S. 103-105).

⁴Wird in (5.4) der Parameter ϱ jeweils durch ϱ_i ersetzt, kann diese Abhängigkeit auch individuenspezifisch modelliert werden.

lich $Z, U_i \stackrel{\text{i.i.d.}}{\sim} N(0, 1)$ vorausgesetzt, resultiert das sogenannte Gauß'sche Einfaktormodell. In diesem sind die X_i standardnormalverteilt und es gilt

$$\text{Corr}[X_i, X_j] = \varrho, \quad i, j = 1, \dots, n, \quad i \neq j.$$

Der Parameter ϱ kann somit als Bonitätskorrelation interpretiert werden. Der Unabhängigkeitsfall ist im Modell (5.4) enthalten, wenn $\varrho = 0$ gilt.

Neben der additiven Verknüpfung von systematischem und idiosynkratischem Faktor Z bzw. U_i sind auch andere Transformationen denkbar. Um speziell einer beschränkten Score-Skala Rechnung zu tragen, könnte z. B. auch folgendes Modell für die Bonitätsvariablen betrachtet werden:

$$X_i = U_i^Z, \quad i = 1, \dots, n. \quad (5.5)$$

Dabei seien die idiosynkratischen Faktoren U_i , $i = 1, \dots, n$ beschränkte Zufallsvariablen, also z. B. beta- oder Vasicek-verteilt⁵, und Z eine positive Zufallsvariable. Ferner seien die U_i , $i = 1, \dots, n$ wiederum i.i.d. und stochastisch unabhängig von Z .

Im Gegensatz zu vielen Quellen⁶ wird für den Zusammenhang zwischen der Ausfallvariable Y und der Bonitätsvariable X an dieser Stelle kein Schwellenwertmodell unterstellt, sondern die Modellgleichung (3.1) durch

$$Y_i | \mathbf{X} = \mathbf{x} \stackrel{\text{st. u.}}{\sim} \text{Ber}(\kappa(x_i)), \quad i = 1, \dots, n \quad (5.6)$$

ersetzt. Die Ausfallvariablen Y_i sind also bedingt auf die Realisation $\mathbf{x} := (x_1, \dots, x_n)$ des Zufallsvektors $\mathbf{X} := (X_1, \dots, X_n)$ aller Bonitäten stochastisch unabhängig (st. u.) und Bernoulli-verteilt mit Parameter $\kappa(x_i)$. Inhaltlich kann das wie folgt interpretiert werden: Das Ausfallverhalten des i -ten Individuums ist nur von der eigenen Bonität abhängig, also unabhängig von der Bonität und dem Ausfallverhalten der anderen Individuen.

Wie schon mehrfach erwähnt wurde, zeigte sich in der empirischen Untersuchung (Kapitel 4), dass die Dempfle-Stute- und die Plug-In-Schätzwerte im Zeitablauf praktisch unkorreliert zur quartalsweisen Ausfallquote waren. Insofern ist man versucht, die zugehörigen Schätzer auch im Modell mit Abhängigkeit zu benutzen. Diese Versuchung wird verstärkt, wenn man bedenkt, dass sich die unbekannten Parameter nur auf das Verhalten von Y_i gegeben X_i auswirken. Die gemeinsame Verteilung des Vektors \mathbf{X} scheint hier nicht von Belang zu sein.

⁵Vgl. zur Beta-Verteilung z. B. Rinne (2008, S. 340-344) und zur Vasicek-Verteilung z. B. Huschens et al. (2010, S. 50-52). Beide Verteilungsklassen sind sehr flexible Wahrscheinlichkeitsverteilungen über dem Intervall $[0, 1]$, die jeweils auch die Gleichverteilung als Spezialfall enthalten.

⁶Vgl. z. B. Höse & Huschens (2011, Abschnitt 3.2), Henking et al. (2006, S. 163), McNeil et al. (2005, Abschnitt 8.3) oder Bluhm et al. (2010, S. 65).

Eine Simulationsstudie, welche auf den Modellgleichungen (5.4) oder (5.5) sowie (5.6) und (3.2) beruht, kann die Vermutung weiter untermauern. Wählt man für die Bonitätsvariablen speziell das Gauß'sche Einfaktormodell, d. h. (5.4) mit $Z, U_i \stackrel{\text{i.i.d.}}{\sim} N(0, 1)$, dann erhält man unter der Annahme $\varrho = 0$ genau das Modell der Simulationsstudie in Ferger & Klotsche (2009). Diese könnte man nun problemlos um extreme, d. h. um sehr kleine oder sehr große und/oder um nur geringfügig unterschiedliche, Risikoniveaus erweitern. Der Unabhängigkeitsfall ist hier jedoch von untergeordnetem Interesse. Eine echte Erweiterung liefert der Fall $\varrho > 0$. Erste Simulationsberechnungen legen die Vermutung nahe, dass die Bruchpunktschätzer in derselben Form wie in Kapitel 3 den unbekannten Bruchpunkt konsistent schätzen. Gleiches gilt für die Plug-In-Schätzer für die Risikoniveaus.⁷

Zum Beweis der Konsistenz kann der Stetigkeitssatz für das argmax-Funktional (argmax-CMT)⁸ von Ferger (2009, S. 25, Satz 4.6) herangezogen werden.⁹ Am Beispiel des Dempfle-Stute-Schätzers soll das entsprechende Vorgehen skizziert werden:

- Identifiziere den Grenzprozess S des empirischen Prozesses S_n .
- Zeige, dass S_n fast sicher gleichmäßig gegen S konvergiert, d. h. dass

$$\|S_n - S\| = \sup_{x \in \mathbb{R}} |S_n(x) - S(x)| \xrightarrow{\text{f. s.}} 0$$

gilt.

- Beweise, dass der Bruchpunkt θ fast sicher eine wohlseparierte Supremalstelle von S ist, d. h. dass

$$\max \{S(\theta), S(\theta-)\} = \sup_{x \in \mathbb{R}} S(x) \text{ f. s.}$$

und

$$\sup_{x \in \mathbb{R}} S(x) > \sup \{S(s) : |s - \theta| > \varepsilon\} \text{ f. s. für alle } \varepsilon > 0.$$

gilt.

⁷Das steht nicht im Widerspruch zu den empirischen Ergebnissen aus Kapitel 4. Dort wird das Zweiklassenmodell behandelt, welches eine sehr stark vereinfachte Abbildung der Realität ist. Stellt man sich eine stetige, monotone Regressionsfunktion vor, werden die Teilausfallquoten gemeinsam mit der Gesamtausfallquote schwanken.

⁸Die Abkürzung CMT steht für *continuous mapping theorem*.

⁹Der Satz ist ohne Beweis auch in Ferger & Klotsche (2009, S. 124, Theorem A.1) zu finden.

Sind die Bedingungen erfüllt, überträgt sich die Konvergenz von S_n gegen S auf die zugehörigen Supremalstellen, d. h.

$$\hat{\theta}_n \xrightarrow{\text{f. s.}} \theta.$$

Bestätigt sich die Vermutung zur Konsistenz, könnten darüber hinaus Überlegungen zur n -Konsistenz oder zur Konstruktion asymptotischer Konfidenzintervalle getroffen werden. Für den Beweis asymptotischer Verteilungsaussagen kommt dabei das Argmax-CMT für Verteilungskonvergenz in Frage.¹⁰

Ferner stellt sich noch die Frage nach der Modellierung zeitlicher Abhängigkeiten und dem Verhalten der Schätzer, wenn Beobachtungen aus unterschiedlichen Perioden zusammengelegt werden. Die Modellierung ist unproblematisch: Ergänze alle Zufallsvariablen, d. h. X_i , Z und U_i , sowie den Stichprobenumfang n in Gleichung (5.4) bzw. (5.5) um einen Zeitindex t und ersetze (5.6) durch

$$Y_{it} | (\mathbf{X}_1, \dots, \mathbf{X}_T) = (\mathbf{x}_1, \dots, \mathbf{x}_T) \stackrel{\text{st. u.}}{\sim} \text{Ber}(\kappa(x_{it})), \quad i = 1, \dots, n_t, \quad t = 1, \dots, T, \quad (5.7)$$

wobei

$$\mathbf{X}_t = (X_{1t}, \dots, X_{n_t t}) \text{ und } \mathbf{x}_t = (x_{1t}, \dots, x_{n_t t}), \quad t = 1, \dots, T.$$

Dieses Modell erlaubt folgende Interpretation: Das Ausfallverhalten Y_{it} des i -ten Individuums zum Zeitpunkt t ist nur von der eigenen Bonität x_{it} zum Zeitpunkt t abhängig. Es ist unabhängig von der eigenen Bonität x_{is} zu allen anderen Zeitpunkten $s \neq t$ sowie von den Bonitäten x_{js} und dem Ausfallverhalten Y_{js} der anderen Individuen $j = 1, \dots, n_s$, $s = 1, \dots, T$. Wird für den stochastischen Prozess (Z_t) z. B. ein stationärer autoregressiver Prozess (AR-Prozess)¹¹ unterstellt, wäre auch Abhängigkeit im Längsschnitt in das Modell integriert. Da die unbekannten Modellparameter nur in der Modellgleichung (5.7) wirken, könnten die bekannten Schätzer hier ebenfalls ihren Dienst tun.

Auch abseits des Kreditrisikos sind Anwendungen der Bruchpunktschätzung denkbar, z. B. bei der Untersuchung des Zusammenhangs zwischen Arbeitslosigkeit und Binnenwanderung. Den Anhaltspunkt dazu geben die Abbildungen 2A bis 2C in Schlömer & Bucher (2001). Unsere Modelle könnten hierfür direkt angewendet werden, wenn positive Wanderungssalden mit 1 und negative mit 0 codiert werden. Abschließend ist daher festzustellen, dass die Methodik der Bruchpunktschätzung auch zur Bearbeitung anderer wirtschaftswissenschaftlicher Fragestellungen vielversprechend ist.

¹⁰Vgl. Ferger (2004, S. 87-88, Theorem 3) oder Ferger (2009, S. 27-28, Satz 4.9).

¹¹Vgl. zum Modell autoregressiver Prozesse z. B. Schlittgen (2012, Abschnitt 3.1), Tsay (2010, Abschnitt 2.4) oder Hamilton (1994, Abschnitt 3.4).

Literaturverzeichnis

- Anderson, R. (2007). *The Credit Scoring Toolkit: Theory and Practice for Retail Credit Risk Management and Decision Automation*. New York: Oxford University Press.
- Bacher, J., Pöge, A. & Wenzig, K. (2010). *Clusteranalyse: Anwendungsorientierte Einführung in Klassifikationsverfahren*. 3. Aufl. München: Oldenbourg.
- Banerjee, M. & McKeague, I. W. (2007). Confidence sets for split points in decision trees. *The Annals of Statistics*, 35(2), S. 543–574.
- BCBS (2006). International convergence of capital measurement and capital standards: A revised framework. <http://www.bis.org/publ/bcbs128.pdf>. Abrufdatum: 11.02.2013.
- Berry, J. D., Dyer, A., Cai, X., Garside, D. B., Ning, H., Thomas, A., Greenland, P., Van Horn, L., Tracy, R. P. & Lloyd-Jones, D. M. (2012). Lifetime risks of cardiovascular disease. *New England Journal of Medicine*, 366(4), S. 321–329. doi: 10.1056/NEJMoa1012848.
- Bishop, Y. M., Fienberg, S. E. & Holland, P. W. (2007). *Discrete Multivariate Analysis. Theory and Applications*. New York: Springer.
- Blöchliger, A. & Leippold, M. (2006). Economic benefit of powerful credit scoring. *Journal of Banking & Finance*, 30(3), S. 851–873.
- Blochwitz, S., Hamerle, A., Hohl, S., Rauhmeier, R. & Rösch, D. (2004). Was leisten Trennschärfemaße für Ratingsysteme? *Zeitschrift für das gesamte Kreditwesen*, 57(22), S. 1275–1278.
- Bluhm, C., Overbeck, L. & Wagner, C. (2010). *Introduction to Credit Risk Modeling*. 2. Aufl. Boca Raton: Chapman & Hall/CRC.
- Blümke, O. (2011). A proposal for a validation methodology for the discriminatory power of a rating system over time. *The Journal of Risk Model Validation*, 5(1), S. 21–44.

- Brockwell, P. J. & Davis, R. A. (1991). *Time Series: Theory and Methods*. 2. Aufl. New York: Springer.
- Bronstein, I. N., Semendjajew, K. A., Musiol, G. & Mühlig, H. (2001). *Taschenbuch der Mathematik*. 5. Aufl. Thun, Frankfurt/M.: Harri Deutsch.
- Calinski, T. & Harabasz, J. (1974). A dendrite method for cluster analysis. *Communications in Statistics*, 3(1), S. 1–27. doi:10.1080/03610927408827101.
- Carey, M. & Hrycay, M. (2001). Parameterizing credit risk models with rating data. *Journal of Banking & Finance*, 25(1), S. 197–270. doi:10.1016/S0378-4266(00)00124-2.
- Casella, G. & Berger, R. L. (2002). *Statistical Inference*. 2. Aufl. Pacific Grove: Duxbury.
- Das, S. R., Duffie, D., Kapadia, N. & Saita, L. (2007). Common failings: How corporate defaults are correlated. *The Journal of Finance*, 62(1), S. 93–117.
- Dempfle, A. & Stute, W. (2002). Nonparametric estimation of a discontinuity in regression. *Statistica Neerlandica*, 56(2), S. 233–242.
- Dudewicz, E. J. & Mishra, S. N. (1988). *Modern mathematical statistics*. New York et al.: Wiley.
- Engelmann, B., Hayden, E. & Tasche, D. (2003). Testing rating accuracy. *Risk*, 16, S. 82–86.
- Everitt, B. S. (1993). *Cluster Analysis*. 3. Aufl. London et al.: Edward Arnold.
- Everitt, B. S. (2006). Mixture distributions – I. In: S. Kotz et al. (Hrsg.). *Encyclopedia of Statistical Sciences*, Bd. 7, S. 4880–4890. Hoboken: John Wiley & Sons, 2. Aufl.
- Everitt, B. S., Landau, S., Leese, M. & Stahl, D. (2011). *Cluster Analysis*. 5. Aufl. Chichester: John Wiley & Sons.
- Fahrmeir, L., Henking, A. & Hüls, R. (2002). Methoden zum Vergleich verschiedener Scoreverfahren am Beispiel der SCHUFA-Scoreverfahren. *RiskNEWS*, 11, S. 20–29.
- Fahrmeir, L., Kneib, T. & Lang, S. (2009). *Regression. Modelle, Methoden und Anwendungen*. 2. Aufl. Berlin, Heidelberg: Springer.
- Ferger, D. (2004). A continuous mapping theorem for the argmax-functional in the non-unique case. *Statistica Neerlandica*, 58(1), S. 83–96.

- Ferger, D. (2009). Stochastische Prozesse mit Strukturbrüchen. Dresdner Schriften zur Mathematischen Stochastik 7/2009. <http://www.math.tu-dresden.de/sto/ferger/DSMS-2009-07.pdf>. Abrufdatum: 08.02.2013.
- Ferger, D. & Klotsche, J. (2009). Estimation of split-points in binary regression. *Statistics & Decisions*, 27(2), S. 93–128. doi:10.1524/stnd.2009.1023.
- Fernandes, J. E. (2005). Corporate credit risk modeling: Quantitative rating system and probability of default estimation. *Techn. Ber.*, Banco BPI. http://www.infofin.net/rating/JEF_CorporateCreditRisk.pdf. Abrufdatum: 08.02.2013.
- Fischer, S. (2002). Erste Umsetzungserfahrungen einer Sparkasse mit dem einheitlichen Rating-Verfahren. *Betriebswirtschaftliche Blätter*, 51, S. 241–249.
- Foglia, A., Iannotti, S. & Marullo Reedtz, P. (2001). The definition of the grading scales in banks' internal rating systems. *Economic Notes. Review of Banking, Finance and Monetary Economics*, 30(3), S. 421–456. doi:10.1111/1468-0300.00066.
- Franke, J., Härdle, W. & Hafner, C. (2004). *Einführung in die Statistik der Finanzmärkte*. 2. Aufl. Berlin et al.: Springer.
- Frey, R., McNeil, A. & Nyfeler, M. (2001). Copulas and credit models. *Risk*, 14(10), S. 111–114.
- Gänssler, P. & Stute, W. (1977). *Wahrscheinlichkeitstheorie*. Berlin: Springer.
- Hamilton, J. D. (1994). *Time Series Analysis*. Princeton: Princeton University Press.
- Hand, D. J. (1997). *Construction and Assessment of Classification Rules*. Chichester: John Wiley & Sons.
- Hansen, B. E. (2008). Uniform convergence rates for kernel estimation with dependent data. *Econometric Theory*, 24(3), S. 726–748. doi:10.1017/S0266466608080304.
- Härdle, W. (1991). *Smoothing Techniques. With Implementation in S*. New York et al.: Springer.
- Härdle, W. (1993). *Applied Nonparametric Regression*. Cambridge: Cambridge University Press.
- Härdle, W., Müller, M., Sperlich, S. & Werwatz, A. (2004). *Nonparametric and Semiparametric Models*. Berlin et al.: Springer.
- Hartung, J., Elpelt, B. & Klösener, K.-H. (2005). *Statistik. Lehr- und Handbuch der angewandten Statistik*. 14. Aufl. München: Oldenbourg.

- Henking, A., Bluhm, C. & Fahrmeir, L. (2006). *Kreditrisikomessung: Statistische Grundlagen, Methoden und Modellierung*. Berlin, Heidelberg: Springer.
- Heuser, H. (1994). *Lehrbuch der Analysis. Teil 1*. 11. Aufl. Stuttgart: Teubner.
- Hinkley, D. V. (1969). Inference about the intersection in two-phase regression. *Biometrika*, 56(3), S. 495–504.
- Hinkley, D. V. (1971). Inference in two-phase regression. *Journal of the American Statistical Association*, 66(336), S. 736–743.
- Hoadley, B. & Oliver, R. M. (1998). Business measures of scorecard benefit. *IMA Journal of Mathematics Applied in Business & Industry*, 9, S. 55–64. doi:10.1093/imaman/9.1.55.
- Höse, B. S. (2007). *Statistische Genauigkeit bei der simultanen Schätzung von Abhängigkeitsstrukturen und Ausfallwahrscheinlichkeiten in Kreditportfolios*. Aachen: Shaker. Zugleich Dissertation, TU Dresden, 2007.
- Höse, S. & Huschens, S. (2003). Sind interne Ratingsysteme im Rahmen von Basel II evaluierbar? Zur Schätzung von Ausfallwahrscheinlichkeiten durch Ausfallquoten. *Zeitschrift für Betriebswirtschaft*, 73(2), S. 139–168.
- Höse, S. & Huschens, S. (2011). Stochastic orders and non-Gaussian risk factor models. *Review of Managerial Science*. doi:10.1007/s11846-011-0071-8.
- Huschens, S. (2010). Kann es Rückzahlungswahrscheinlichkeiten von 100% geben? Dresdner Beiträge zu Quantitativen Verfahren 53/10. <http://www.tu-dresden.de/wwqvs/publ/DBQV53-10.pdf>. Abrufdatum: 08.02.2013.
- Huschens, S., Lehmann, C. & Tillich, D. (2010). Sensitivities and worst-case correlations for hitting probabilities of portfolio tranches. *The Journal of Risk Model Validation*, 4(1), S. 49–69.
- Huschens, S. & Locarek-Junge, H. (2002). Konzeptionelle und statistische Grundlagen der portfolioorientierten Kreditrisikomessung. In: A. Oehler (Hrsg.). *Kreditrisikomanagement. Kernbereiche, Aufsicht und Entwicklungstendenzen*, S. 89–114. Stuttgart: Schäffer-Poeschel, 2. Aufl.
- Karr, A. F. (1993). *Probability*. New York et al.: Springer.
- Korczak, D. & Wilken, M. (2009). Verbraucherinformation Scoring. <http://www.bmelv.de/cae/servlet/contentblob/638114/publicationFile/36036/Scoring.pdf>. Abrufdatum: 08.02.2013.

- Kramer, F. & Löffler, G. (2010). Corporate bond defaults are consistent with conditional independence. *Journal of Credit Risk*, 6(2), S. 3–35. doi:10.1016/j.jfi.2010.03.002.
- Krämer, W. & Bucker, M. (2011). Probleme des Qualitätsvergleichs von Kreditausfallprognosen. *AStA – Wirtschafts- und Sozialstatistisches Archiv*, 5(1), S. 39–58. doi:10.1007/s11943-011-0096-0.
- Kreiß, J.-P. & Neuhaus, G. (2006). *Einführung in die Zeitreihenanalyse*. Berlin et al.: Springer.
- Krink, T., Paterlini, S. & Resti, A. (2008). The optimal structure of PD buckets. *Journal of Banking & Finance*, 32(10), S. 2275–2286. doi:10.1016/j.jbankfin.2007.12.042.
- Kuhlish, W. (2003). Estimation of a threshold-value in the context of air pollution and health. *Economic Quality Control*, 18(2), S. 241–249. doi:10.1515/EQC.2003.241.
- Lando, D. & Nielsen, M. S. (2010). Correlation in corporate defaults: Contagion or conditional independence? *Journal of Financial Intermediation*, 19(3), S. 355–372.
- Lehmann, C. (2013). *Sektorale Unterschiede in Kreditausfalldaten: Modellierung und statistische Inferenz basierend auf Bonitätsscores*. Dresden: TUDpress. Zugleich Dissertation, TU Dresden, 2012.
- Lehmann, E. L. (1999). *Elements of Large-Sample Theory*. New York: Springer.
- Lehmann, E. L. & Casella, G. (1998). *Theory of Point Estimation*. 2. Aufl. New York: Springer.
- Li, Q. & Racine, J. S. (2007). *Nonparametric Econometrics - Theory and Practice*. Princeton, Oxford: Princeton University Press.
- Maulik, U. & Bandyopadhyay, S. (2002). Performance evaluation of some clustering algorithms and validity indices. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(12), S. 1650–1654. doi:10.1109/TPAMI.2002.1114856.
- Mays, E. (2004a). Scorecard performance measures. In: E. Mays (Hrsg.). *Credit Scoring for Risk Managers: The Handbook for Lenders*, Kap. 6, S. 105–120. Mason: Thomson - South Western.
- Mays, E. (2004b). Understanding the KS-statistic. In: E. Mays (Hrsg.). *Credit Scoring for Risk Managers: The Handbook for Lenders*, Kap. 7, S. 121–130. Mason: Thomson - South Western.

- McCullagh, P. & Nelder, J. (1999). *Generalized Linear Models*. 2. Aufl. Boca Raton et al.: Chapman & Hall/CRC.
- McNeil, A. J., Frey, R. & Embrechts, P. (2005). *Quantitative Risk Management: Concepts, Techniques and Tools*. Princeton: Princeton University Press.
- Meintrup, D. & Schäffler, S. (2005). *Stochastik – Theorie und Anwendungen*. Berlin, Heidelberg: Springer.
- Milligan, G. W. & Cooper, M. C. (1985). An examination of procedures for determining the number of clusters in a data set. *Psychometrika*, 50(2), S. 159–179.
- Moody's Investors Service (2012). Annual default study: Corporate default and recovery rates, 1920-2011 - excel data. <http://www.moodys.com/researchandratings/research-type/default-ratings-analytics/rating-performance/003009006/4294964666%204294962002/4294966848/0/0/-/0/-/-/-/-/-/-/-/de/ger/pdf/rra>. Abrufdatum: 08.02.2013.
- Mosler, K. & Schmid, F. (2011). *Wahrscheinlichkeitsrechnung und schließende Statistik*. 4. Aufl. Berlin, Heidelberg: Springer.
- Müller, P. H., Hrsg. (1991). *Wahrscheinlichkeitsrechnung und mathematische Statistik: Lexikon der Stochastik*. 5. Aufl. Berlin: Akademie Verlag.
- Oesterreichische Nationalbank & Finanzmarktaufsicht (2004). Ratingmodelle und -validierung. Leitfadenreihe zum Kreditrisiko, Wien. http://www.oenb.at/de/img/leitfadenreihe_ratingmodelle_tcm14-11172.pdf. Abrufdatum: 08.02.2013.
- Parzen, E. (1962). On estimation of a probability density function and mode. *The Annals of Mathematical Statistics*, 33(3), S. 1065–1076.
- Pfingsten, A. & Schröck, G. (2000). Bedeutung und Methodik von Krediteinstufungsmodellen im Bankwesen. In: A. Oehler (Hrsg.). *Kreditrisikomanagement - Portfolio-modelle und Derivate*, S. 1–23. Stuttgart: Schäffer-Poeschel.
- Price, K. V., Storn, R. & Lampinen, J. A. (2005). *Differential Evolution. A Practical Approach to Global Optimization*. Berlin et al.: Springer.
- Rice, J. A. (2007). *Mathematical Statistics and Data Analysis*. 3. Aufl. Belmont: Brooks/Cole.
- Rinne, H. (2008). *Taschenbuch der Statistik*. 4. Aufl. Frankfurt/M.: Harri Deutsch.
- Ross, S. M. (2006). *Statistik für Ingenieure und Naturwissenschaftler*. 3. Aufl. München: Elsevier.

- Schlittgen, R. (2012). *Angewandte Zeitreihenanalyse mit R*. München: Oldenbourg.
- Schlömer, C. & Bucher, H. (2001). Arbeitslosigkeit und Binnenwanderungen. Auf der Suche nach einem theoriegestützten Zusammenhang. *Informationen zur Raumentwicklung*, 1, S. 33–47.
- Schmidt, C., Herting, B., Prieur, S., Junghanns, S., Schweitzer, K., Globas, C., Schöls, L., Antoni, S., Ferger, D., Reichmann, H., Wilhelm, H., Berg, D. & Ziemssen, T. (2007). Pupil diameter in darkness differentiates progressive supranuclear palsy (psp) from other extrapyramidal syndromes. *Movement Disorders*, 22(14), S. 2123–2126. doi:10.1002/mds.21721.
- Schuster, E. F. (1969). Estimation of a probability density function and its derivatives. *The Annals of Mathematical Statistics*, 40(4), S. 1187–1195.
- Serfling, R. J. (1980). *Approximation Theorems of Mathematical Statistics*. New York et al.: John Wiley & Sons.
- Shorack, G. R. (2000). *Probability for Statisticians*. New York: Springer.
- Silverman, B. W. (1978). Weak and strong uniform consistence of the kernel estimate of a density and its derivatives. *The Annals of Statistics*, 6(1), S. 177–184.
- Silverman, B. W. (1994). *Density Estimation for Statistics and Data Analysis*. London: Chapman & Hall.
- Sobehart, J. & Keenan, S. (2001). Measuring default accurately. *Risk*, 15(3), S. S31–S33.
- Steland, A. (2012). *Financial Statistics and Mathematical Finance - Methods, Models and Applications*. Chichester: John Wiley & Sons.
- Stocks, T., Rapp, K., Bjørge, T., Manjer, J., Ulmer, H., Selmer, R., Lukanova, A., Johansen, D., Concin, H., Tretli, S., Hallmans, G., Jonsson, H. & Stattin, P. (2009). Blood glucose and risk of incident and fatal cancer in the metabolic syndrome and cancer project (Me-Can): Analysis of six prospective cohorts. *PLoS Med*, 6(12), S. 1–14. doi:10.1371/journal.pmed.1000201.
- Storn, R. & Price, K. (1997). Differential evolution – a simple and efficient heuristic for global optimization over continuous spaces. *Journal of Global Optimization*, 11(4), S. 341–359. doi:10.1023/A:1008202821328.
- Titterton, D. M. (2006). Mixture distributions – II. In: S. Kotz et al. (Hrsg.). *Encyclopedia of Statistical Sciences*, Bd. 7, S. 4890–4899. Hoboken: John Wiley & Sons, 2. Aufl.

- Toutenburg, H. (2003). *Lineare Modelle*. 2. Aufl. Heidelberg: Physica.
- Tsay, R. S. (2010). *Analysis of Financial Time Series*. 3. Aufl. Hoboken: John Wiley & Sons.
- Tutz, G. (2000). *Die Analyse kategorialer Daten: Anwendungsorientierte Einführung in Logit-Modellierung und kategoriale Regression*. München, Wien: Oldenbourg.
- van der Vaart, A. W. (1998). *Asymptotic Statistics*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge: Cambridge University Press.
- Vendramin, L., Campello, R. J. G. B. & Hruschka, E. R. (2009). On the comparison of relative clustering validity criteria. In: *Proceedings of the Ninth SIAM International Conference on Data Mining*, S. 733–744. http://www.siam.org/proceedings/datamining/2009/dm09_067_vendramin1.pdf. Abrufdatum: 08.02.2013.
- Wasserman, L. (2006). *All of Nonparametric Statistics*. New York: Springer.
- Wehberg, S., Sauerbrei, W. & Schumacher, M. (2008). Diagnoseverfahren: Wertigkeit der Sonographie bei der Differenzierung von gut- und bösartigen Brusttumoren bei Patientinnen mit klinischen Symptomen. In: M. Schumacher & G. Schulgen (Hrsg.). *Methodik klinischer Studien. Methodische Grundlagen der Planung, Durchführung und Auswertung*, Kap. 18, S. 319–340. Berlin, Heidelberg: Springer, 3. Aufl. doi: 10.1007/978-3-540-85136-3-18.
- White, H. (2001). *Asymptotic Theory for Econometricians*. San Diego et al.: Academic Press.
- Wied, D. & Weißbach, R. (2012). Consistency of the kernel density estimator - a survey. *Statistical Papers*, 53(1), S. 1–21. doi:10.1007/s00362-010-0338-1.
- Witting, H. (1985). *Mathematische Statistik I. Parametrische Verfahren bei festem Stichprobenumfang*. Stuttgart: Teubner.
- Zinde-Walsh, V. (2008). Kernel estimation when density may not exist. *Econometric Theory*, 24(3), S. 696–725. doi:10.1017/S0266466608080298.